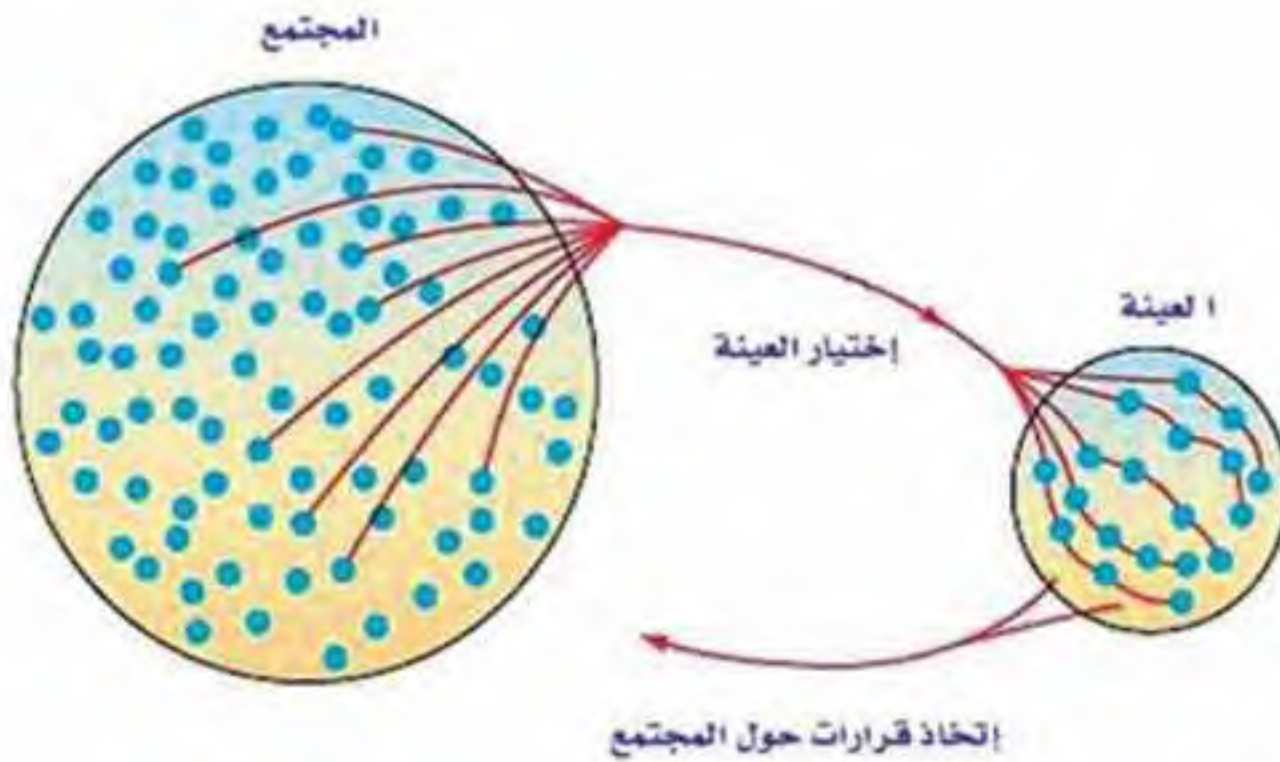
د. حسين علوان مطلق

جمع البيانات وطرق المعاينة





جامعة الملك فهد للبترول والمعادن
عمادة البحث العلمي

العبيكان
obeikan

جمع البيانات
وطرق المعاينة

تأليف

د. حسين علوان مطلق

أستاذ في قسم الرياضيات والإحصاء — كلية العلوم

جامعة الملك فهد للبترول والمعادن

العبيكان
Obeikan



جامعة الملك فهد للبترول والمعادن

عمادة البحث العلمي

ح مكتبة العبيكان، 1430هـ

فهرسة مكتبة الملك فهد الوطنية أثناء النشر

الحميدان، سليمان صالح

جمع البيانات وطرق المعاينة / حسين علوان مطلق – ط 1 – الرياض، 1430هـ

202ص، 16.5×24سم

ردمك: 978-9960-؟؟-؟؟-؟

1- .... أ. العنوان

ديوي: 658 1430/2291

رقم الإيداع: 1430/2291

ردمك: 978-9960-؟؟-؟؟-؟



صدر هذا الكتاب بدعم من جامعة الملك فهد للبترول والمعادن ضمن اتفاقية نشر خاصة بين شركة العبيكان للأبحاث والتطوير وعمادة البحث العلمي في الجامعة

الطبعة الأولى

1430هـ/2009م

الناشر: العبيكان للنشر

الرياض – شارع العليا العام – جنوب برج المملكة

هاتف: 2937574/ 2937581 فاكس: 2937588

ص.ب: 67622 الرمز 11517

التوزيع: مكتبة العبيكان

الرياض – العليا – تقاطع طريق الملك فهد مع العروبة

هاتف: 4160018/4654424 فاكس: 4650129

ص.ب: 62807 الرمز 11595

بسم الله الرحمن الرحيم

المحتويات

الفصل الثالث: متطلبات العينة الجيدة

الفصل الرابع: تنفيذ المساحة بالعينة

الفصل الخامس: العينة العشوائية البسيطة

الفصل السادس: العينة العشوائية الطبقية

الفصل السابع: التقدير بطريقة النسبة والانحدار والفرق

الفصل الثامن: العينة العشوائية المنتظمة

الفصل التاسع: العينة العنقودية بمرحلة واحدة

الفصل العاشر: العينة العنقودية بمرحلتين

الفصل الحادي عشر: المعاينة مختلفة الاحتمالات

الفصل الثاني عشر: المعاينة المزدوجة

الفصل الثالث عشر: تقدير حجم المجتمع

الفصل الرابع عشر: المعاينة بطريقة الإمساك وإعادة الإمساك

الفصل الخامس عشر: المعاينة بالخط القاطع

الفصل السادس عشر: المعاينة بخط التقاطع

الفصل السابع عشر: المعاينة المكانية أو كريج انج

الفصل الثامن عشر: المعاينة العنقودية المكيفة

الفصل التاسع عشر: المعاينة الشبكية

الفصل العشرون: معاينة المجموعات المرتبة

الفصل الحادي والعشرون: طريقة السكين الحادة لإعادة المعاينة

الفصل الثاني والعشرون: طريقة السكين الحادة لإعادة المعاينة

الفصل الثالث والعشرون: أخطاء عدم الإجابة

الفصل الرابع والعشرون: طرق أخرى للمعاينة

الملاحق

بين يدي الكتاب

الحمد للّه الذي بنعمته تتم الصالحات، والصلاة والسلام على رسول اللّه صلى اللّه عليه وسلم وعلى آله وصحبه ومن اتبع سنته إلى يوم الدين. سبحانك لا علم لنا إلا ما علمتنا إنك أنت العليم الحكيم. وبعد:

فإنّ موضوع هذا الكتابِ وهو جمع البيانات وطرق المعاينة يُعَدُّ من الموضوعات المهمة جداً في حياتنا اليومية، حيث أصبح العالم منشغلاً بجمع وتحليل البيانات حول جميع نواحي الحياة اليومية، خصوصاً السياسية والاجتماعية والاقتصادية...إلخ، فضلاً عن الجوانب المتعلقة بالبحث العلمي من فروع المعرفة المختلفة، كالكيمياء والفيزياء والأحياء...إلخ

توجد هنالك طريقتان لجمع المعلومات أو البيانات من المجتمعات المختلفة: إما المسح الشامل لجميع عناصر المجتمع، أو سحب عينة من المجتمع ودراستها ومن ثم تعميم النتائج التي حصلنا عليها من العينة على المجتمع، وما لم تكن العينة سُحبت من المجتمع وفق الطرق الإحصائية العلمية، وهو ما يسمى بالعينة العشوائية فإننا لا نستطيع الاستفادة من النتائج التي حصلنا عليها. وبما أن عملية المسح الشامل لجميع عناصر المجتمع أصبحت في أغلب الأحوال غير ممكنة إن لم تكن مستحيلة؛ لذلك أصبح جمع البيانات باستخدام العينات العشوائية يتبوأ مكانة كبيرة.

لقد تم تأليف هذا الكتاب وفي البال أن يخدم الباحث العربي المتخصص في مختلف فروع المعرفة أو طلاب الدراسات العليا، الذين يهمهم أن يستخدموا طريقة أو أكثر من طرق المعاينة ليقوموا بتطبيقها لجمع البيانات وفق الشروط العلمية الصحيحة، فسيجدون هذا الكتاب خير معين لهم بإذن اللّه.

يمكن أن تستخدم الفصول الثلاثة عشر الأولى لتكون عبارة عن أول مساق دراسي لطلبة قسم الإحصاء في مجال العينات. كذلك يمكن استخدام الفصول العشرة الأخيرة لتكون مساقاً ثانياً لطلبة قسم الإحصاء في مجال العينات، ويستطيع المدرس أن يختار أي فصل من الفصول الأخيرة لتكون ضمن المساق الأول على أن يحتوي المساق على الفصول التسعة الأولى التي تشكل طرق المعاينة الأساسية وهي التي يتعين على طالب الإحصاء دراستها.

ولكي تتم الفائدة القصوى من الكتاب لا بد أن يكون لدى القارئ معلومات عن مبادئ الإحصاء، أو يكون لديه على الأقل مساق واحد في مقدمة الإحصاء.

قبل أن أنهي كلامي لا بد أن أتقدم بالشكر والعرفان لجامعة الملك فهد للبترول والمعادن على دعمها لإنجاز هذا الكتاب. كذلك لا يفوتني أن أشكر الدكتور عبدالله عمر الحاج إبراهيم الذي قام جزاه الله خيراً بمراجعة الكتاب والتأكد من سلامته من الأخطاء اللغوية، والدكتور وليد الصباح الذي قام مشكوراً بمراجعة الكتاب، وأبدى بعض الملاحظات والاقتراحات التي أخذنا بها وجزاه الله خيراً.

الفصل الأول

المقدمة
Introduction

1.1 الفكرة الأساسية للمعاينة والتقدير

عندما نبدأ بتنفيذ المعاينة الأساسية، يكون المجتمع محتوياً على عدد معروف ومحدد N من الوحدات مثل مجموعة من الناس أو قطع من الأرض أو عدد من المنازل وغيرها. مع كل وحدة هنالك قيمة مصاحبة لمتغير نرغب في دراسته. وهذه القيم ينظر إليها على أنها ثابتة، وليست متغيراً عشوائياً. وترقم وحدات المجتمع على النحو التالي1,2, □,N.

نسحب عينة من وحدات المجتمع ونقوم بمشاهدة وحدات العينة فقط. تحتوي البيانات على قيمة المتغير، ومعلومات أخرى حول الوحدة، فعلى سبيل المثال كل بئر نفط يجري حفره نقوم بتسجيل مقدار النفط في هذا البئر، بالإضافة إلي رقم البئر وموقعه وأي معلومات أخرى حوله. كذلك نقوم بحفظ معلومات حول متغيرات إضافية مثل عمق البئر، ونوع الخام وغيرها. أما في المسوحات التي تجرى على النباتات، فمن المتغيرات الإضافية وجود نباتات أخرى، وكثافة وجودها في المنطقة. أما في حالة استطلاعات الرأي فمن المتغيرات الإضافية التي غالباً ما نسجلها الجنس والعمر ومقدار الدخل وغيرها.

تسمى الطريقة التي نسحب بها وحدات العينة من المجتمع بتصميم المعاينة. مع معظم تصاميم المعاينة المعروفة، يُحدد التصميم بواسطة إعطائه لكل عينة S احتمالاً مقداره P(s) لاختيار تلك العينة. فعلى سبيل المثال في المعاينة العشوائية البسيطة وبحجم n وحدة، فإن الاحتمال P(s) يكون متساوياً لجميع العيناتs .

إن الاستنتاجات المعتادة من المعاينة هي تقدير بعض صفات (معلمات) المجتمع مثل المتوسط، أو المجموع الكلي، أو التباين، أو النسبة بعد مشاهدة العينة فقط. بالإضافة إلى ذلك في معظم حالات المعاينة والتقدير ربما نرغب بتقييم دقة التقديرات التي أوجدنها لبعض معلمات المجتمع باستخدام الثقة المصاحبة للتقدير، وفي معظم الحالات نستخدم فترة الثقة.

من وجهة النظر الأساسية للمعاينة، إذا كبرنا حجم العينة إلى أن تكون مساوية لحجم المجتمع N، أي تم مشاهدة جميع وحدات المجتمع، فإن صفات أو معلمات المجتمع ستكون معلومة تماماً، وإن الغموض في التقديرات التي أوجدناها باستخدام العينات يعود سببه إلى أن جزءاً فقط من المجتمع تم مشاهدته باستخدام العينة، بينما تبقى صفات المجتمع ثابتة، يعتمد التقدير على أي عينة من المجتمع جرى سحبها، إذا كان لكل عينة ممكنة يبقى التقدير قريباً من القيمة الحقيقية لصفة المجتمع، ويوجد غموض صغير في إستراتيجية المعاينة؛ لذا فإن هذا النوع من إستراتيجيات المعاينة مطلوب، ومن الناحية الأخرى إذا كانت قيمة التقدير تختلف بشكل كبير من عينة إلى أخرى، سيكون الغموض المصاحب لهذه الطريقة كبيراً.

يمكننا أن نحصل على تقديرات غير متحيزة إلى صفات المجتمع مثل المتوسط أو المجموع الكلي للمجتمع إذا تم اختيار تصميم المعاينة بعناية، وكذلك إذا تم اختيار طريقة التقدير الملائمة. أن يكون التقدير غير متحيز

يعني أن معدل جميع العينات التي يمكن سحبها يساوي القيمة الحقيقية لصفة المجتمع التي يجري تقديرها؛ لذا من خلال تصميم المعاينة وطريقة التقدير يمكننا أن نحصل على تقدير غير متحيز إلى كثافة الطيور في منطقة دراسية معينة بغض النظر فيما إذا كان توزيع الطيور في المجتمع بشكل متساوٍ أو إذا كانت متجمعة في مجموعات قليلة في منطقة الدراسة.

إن استخدام تصميم المعاينة الاحتمالية مثل المعاينة العشوائية البسيطة؛ من شأنه أن يوفر لنا تقديرات غير متحيزة إلى متوسط وإلى المجموع الكلي للمجتمع، كما يوفر لنا تقديرًا غير متحيز إلى التباين، الذي يستخدم للتأكد من موثوقية نتائج المسح بالعينة، كذلك يمكننا أن نحصل على تقديرات غير متحيزة من التصاميم التي تكون الاحتمالات فيها غير متساوية.

بالإضافة إلى هدف الحصول على تقدير غير متحيز من المسح بالعينة، هنالك هدف الدقة أو ما يسمي بالتباين القليل للتقديرات، كذلك نهدف للحصول على تصميم يكون سهل الاستخدام وتكاليف تنفيذه قليلة. لقد أدت هذه الأهداف إلى تطوير تصاميم معاينة كثيرة ومختلفة منها المعاينة العشوائية، ومعاينة الاحتمالات المختلفة، استخدام المعلومات الإضافية، والطبقية، والمنتظمة، والعنقودية وغيرها من تصاميم المعاينة المختلفة التي سنتناول الكثير منها في هذا الكتاب.

21 وحدات المعاينة

قد يسهل تحديد وحدات المعاينة ومن ثم وضع كشف بجميع أسماء الوحدات وهو ما يسمى بإطار المعاينة في كثيرٍ من المجتمعات مثل الإنسانية والمعاهد والمنازل وغيرها. بغض النظر عن الصعوبات العملية في تجهيز إطار المعاينة. تكون الوحدات في هذه المجتمعات إما شخصاً، أو منزلاً، أو مستشفى، أو شركة، أو غيرها. ويكون هناك كشف كامل لجميع

الأشخاص، أو المنازل، أو المستشفيات، أو الشركات في المجتمع الذي نرغب في دراسته ليعطينا إطار المعاينة المثالي، وغالباً ما تعتري عملية بناء إطار المعاينة صعوبات كثيرة عند القيام بتنفيذ ذلك ميدانياً، على أن استخدام دليل الهاتف إطارًا للمعاينة مثلاً في المسوحات التي تستخدم الهاتف، تكتنفه بعض المشكلات، منها أنه ليس جميع أفراد المجتمع لديهم هواتف، كذلك هنالك أشخاص أسماؤهم ليست موجودة في دليل الهاتف بناء على طلبهم؛ لذا فإن عينة عشوائية باستخدام دليل الهاتف إطارًا للمعاينة لا يعطي فرصة للعوائل التي لا يوجد لديها هواتف أو أن هواتفها غير موجودة في الدليل بأن تكون في العينة، أما في حالة المسوحات الخاصة بالشركات فقد تكون الكشوفات ناقصة، أو قديمة ولم يجر تحديثها لتشمل الشركات التي أسست بعد إعداد الكشف.

هنالك مجتمعات كثيرة غير واضح ماذا ستكون وحدة المعاينة فيها؟ في المسوحات الخاصة بالمواد الطبيعية، أو المحاصيل الزراعية في منطقة الدراسة، يمكن أن تكون وحدة المعاينة عبارة عن قطعة أرض بمساحة معينة، يتم الحصول عليها من خلال تقسيم المنطقة إلى مجموعة من القطع، على الخارطة ومن ثم نقوم بسحب عينة من القطع، ولكن إذا كان الشخص حراً في اختيار حجم القطع وأشكالها، فمثل هذا الاختيار ربما يؤدي إلى التأثير على تكاليف المسح بالعينة وعلى دقة المقدرات، أما في حالة طرق المعاينة التي تعتمد على اختيار نقطة بطريقة عشوائية في منطقة الدراسة، وتتمحور وحدة المعاينة حولها، فإنه من المحتمل أن تتقاطع وحدات العينة مع بعضها، وهذا سيؤدي إلى أن تكون عدد وحدات المجتمع غير محدودة.

أما في المجتمعات التي تتصف بعدم الاستقرار أو المراوغة مثل مجتمع الحيوانات البرية والبحرية، فإن مشكلة اكتشاف الوحدة يؤدي دوراً مهماً،

وهنالك دالة لاكتشاف الوحدات، لذا فإن دور الوحدة سيتأثر كثيراً بدالة الاكتشاف التي تكون مصاحبة للطريقة التي سيشاهد بها المجتمع، وتحديد المواضع للقيام بمشاهدة الوحدات.

قد يكون المتغير الذي نرغب في دراسته يختلف من منطقة إلى أخرى بشكل مستمر ضمن منطقة الدراسة، فعلى سبيل المثال في حالة المسوحات الخاصة بالبحث عن النفط في منطقة دراسية ما، قد يكون المتغير هو عمق أو كمية النفط في الموضع. المتغيران هنا ليس بالضرورة أن يكونا مصاحبين لمجموعة من الوحدات المحدودة في المنطقة، ولكن بالأحرى يقاسا أو يقدرا لنقطة معينة، أو المجموع لمنطقة جزئية من المجتمع بأي حجم أو شكل.

على الرغم من أن بعض طرق المعاينة التي سنتناولها في هذا الكتاب تتجاوز مسألة إطار المعاينة الذي يقسم المجتمع إلى مجموعة من الوحدات المحددة، ومن ثم نقوم بسحب العينة من هذه المجموعة من الوحدات، إلا أنَّ بعض صفات طرق المعاينة المعروفة مثل المعاينة العشوائية البسيطة، والمعاينة الطبقية مثل عدم تحيز التقدير وتقديرات النسبة وغيرها ستبقى تستعمل.

3.1 أخطاء المعاينة وأخطاء أخرى

تفترض طرق المعاينة الأساسية أن المتغير الذي نرغب بدراسته يمكن قياسه في جميع وحدات العينة وبدون أخطاء، ومن ثم فإن الخطأ في التقدير يكون سببه أن جزءاً من المجتمع فقط جرى سحبه بالعينة، وهذا الخطأ يُسمى خطأ المعاينة (sampling errors)، ولكن في الحياة العملية وفي حالات المسوحات الحقيقية هنالك أخطاء أخرى (Nonsampling errors) تحدث ليس سببها المعاينة وإنما أسباب لا علاقة لها بالمعاينة، قد لا يرد أحد على الهاتف في حالة المسوحات التي تجرى بالهاتف وذلك لكونهم خارج المنزل، أو يرفضون الإجابة عن بعض الأسئلة الموجودة ضمن الاستبانة لسبب أو لآخر. ربما يكون

هؤلاء الرافضون غير نموذجيين في المجتمع، وهذا سيؤدي لأن تكون العينة غير ممثلة للمجتمع ومن ثم ستكون التقديرات متحيزة في المسوحات الخاصة بالسمك، ربما بعض المواضع التي جرى سحبها في العينة لم يتم مشاهدتها لكون أحوال الطقس سيئة وقت إجراء المسح، أو تكون بعيدة جداً من الشاطئ بحيث يصعب الوصول إليها باستخدام الوسائل المتاحة للمسح.

إن مشكلة عدم الإجابة من المشكلات التي تعتري كثيراً المسوحات، التي مقياسها يكون قلة عدد المستجيبين لبعض الأسئلة الموجودة في الاستبانة، أو رفض الإجابة عن جميع أسئلة الاستبانة لأسباب كثيرة منها: أن الأشخاص لا يعيرون أهمية لمثل هذه الأمور، أو لكونهم مشغولين، أو لكون بعض الأسئلة تتعلق بأمور شخصية لا يحبون الإفصاح عنها للآخرين. يمكننا أن نقلل تأثير عدم الإجابة وذلك من خلال جهود أكبر من قبل الباحثين أو العادّين منها: زيارة الوحدة مرة أخرى أو عدة مرات، أو إعادة صياغة بعض الأسئلة المحرجة التي تتعلق بالأمور الشخصية، أو استخدام بعض المعلومات أو المتغيرات الإضافية، أو بناء نماذج رياضية خاصة بعدم الإجابة وغيرها.

هنالك أخطاء أخرى مثل أخطاء القياس أو التسجيل التي يمكن أن تحدث لسبب أو لآخر منها أن أجهزة القياس غير دقيقة، أو قديمة، أو لا يعرف العادون كيفية استخدامها بشكل فاعل، لابد لنا من محاولة السيطرة على هذا النوع من الأخطاء قدر المستطاع، هنالك حالات يمكننا أن نبني فيها نموذجاً لأخطاء القياس بصورة مستقلة عن أخطاء المعاينة، لأخطاء عدم الإجابة التي تحدث في المجتمعات البرية والمجتمعات المراوغة، في مسحات الطيور من المألوف ألاّ يستطيع العاد أن يكتشف جميع الطيور الموجودة بجواره في إحدى وحدات المعاينة، في مسوحات الأسماك لا يمكن صيد جميع الأسماك التي تقع في طريق الشبكة، وليس جميع المشردين الذين لا يجدون

سكناً خاصاً بهم يمكن الوصول إليهم، سنتناول في هذا الكتاب مجموعة من الطرق التي تتعامل مع هذا النوع من المجتمعات منها: طريقة الخط القاطع، وطريقة خط التقاطع، وطريقة الإمساك وإعادة الإمساك وغيرها.

4.1 لمحات من تاريخ المعاينة

إن أول من استخدم المعاينة هو ملك فارس وذلك في عام 440 قبل الميلاد حسب المعلومات المتوافرة لدينا. ولو أن البابليين قد استخدموا نوعاً من المعاينة لتقدير منتجاتهم من القمح والمحاصيل الأخرى التي تُعَد المادة الغذائية الرئيسة لساكني بلاد ما بين النهرين وذلك منذ 3000 سنة من قبل الميلاد. أما الملك الفارسي فقد طلب أن يحصى جيشه الجرار الذي استخدمه لاحتلال بلاد اليونان، والطريقة التي اعتمدت لتقدير أعداد الجند تتلخص في أنه جرى تحديد منطقة معينة وطلب من مجموعة من الجند الوقوف مصطفّين بعضهم جنب بعض، وتم حصر عدد الجنود الذين يمكن أن تحتويهم هذه القطعة، ومن ثم وضع حولها سياجاً، وطلب من جميع الجند المرور من خلال هذه المنطقة، وتم مشاهدة عدد المرات التي امتلأت بها، ومن ثم تم تقدير عدد الجنود من خلال ضرب سعة المنطقة، في عدد المرات التي امتلأت بها هذه المنطقة، وقد تم تقدير عدد الجند بنحو 1,700,000 جندي، لم تذكر لنا كتب التأريخ كيف تم اختيار الجنود الذين استخدموا في المرة الأولى لمعرفة عدد الجنود الذين يستطيعون الوقوف بهذه المنطقة، وكذلك لم نعرف إذا تمت عملية التأكد من دقة التقدير أم لا.

لقد طُورت معظم طرق المعاينة التي نعرفها اليوم في القرن العشرين. كان هنالك نقاش بين مستخدمي طرق المعاينة مع بدايات القرن العشرين حول جدارة وكفاءة العينة العشوائية بالمقارنة بالعينة التي يجري اختيار وحداتها من قبل المستخدم لتكون ممثلة للمجتمع قدر الإمكان. لقد تم تطوير طريقة

المعاينة العشوائية البسيطة خلال العقدين الأولين من القرن العشرين. لقد قام Neyman (1934) بالمقارنة بين الطريقتين، ووضع المفاهيم الأساسية للمعاينة الاحتمالية التي نستطيع من خلالها سحب عينة عشوائية من توزيع احتمالي محدد، لقد طُورت معظم طرق المعاينة المعروفة قبل نهايات العقد الثالث من القرن العشرين، ومن هذه الطرق: المعاينة الطبقية، والمعاينة المنتظمة، والمعاينة العنقودية، والمعاينة متعددة المراحل، والمعاينة المزدوجة، والمعاينة متعددة الوجوه. لقد استخدمت دائرة الإحصاءات الأمريكية المعاينة الاحتمالية لأول مرة في إحصاءات العاطلين عن العمل مع بداية عام 1940م. أما معاينات الاحتمالات غير المتساوية فقد طُورت في أربعينيات وخمسينيات القرن العشرين.

لقد استمرت نظريات وطرق المعاينة في التطور حتى نهايات القرن العشرين وحتى يومنا هذا، قام بعض الباحثين من بينهم: Godambe في خمسينيات القرن العشرين بدراسة الجوانب النظرية من طرق المعاينة، مما ساعد كثيراً في توضيح قضايا الاستدلال في المعاينة، ومهد الطريق لتقديم طرق جديدة للمعاينة، لقد جرى تطوير مجموعة من التصاميم وطرق الاستدلال استجابة لبعض المشكلات الصعبة التي واجهت الباحثين في مختلف فروع المعرفة، أسهمت بتطوير طرق المعاينة. كذلك الاختلاف بين وجهات النظر حول معاينة تعتمد على التصميم، ومعاينة تعتمد على النموذج، أسهمت في تطوير طرق للجمع بين الطريقتين، التطورات الأخيرة حول تحليل البيانات مع فقدان بعضها أسهم في تطوير طرق تحليلية جديدة أُستفيد منها في أوجه متعددة.

5.1 تصاميم وخلاصة المعاينة

يؤدي تصميم العينة دوراً مهماً في تحديد دقة التقديرات؛ لذا فإن الطريقة التي تسحب بها العينة مهمة كما للشكل الرياضي للمقدر من الأهمية.

يتضمن تصميم المعاينة: خطة سحب العينة، وطريقة التقدير. كلٌّ منهما يحتوي على مجموعة من المعالم: أولاً لابد أن نعرف وحدات المعاينة (الأولية والثانوية وهكذا)ثُمّ نختار إحدى طرق المعاينة مثل (المعاينة العشوائية البسيطة، أو المعاينة الطبقية، أو المعاينة المنتظمة، وهكذا)؛ نوزع الوحدات بين الطبقات المختلفة بطريقة تضمن التوزيع الأمثل لهذه الوحدات بين الطبقات، ومعاينة باحتمالات مختلفة.

نحاول تقدير معالم المجتمع باستخدام التقديرات الخطية، يمكننا أن نستخدم معلومات إضافية لتقديرات النسبة والانحدار، نحول المقدرات غير الخطية (على سبيل المثال تقدير النسبة) إلى مقدرات خطية قبل حساب التباين. كما أنه لابد من بذل جهد مهم في مسوحات المعاينة للسيطرة أو تقليل أخطاء عدم الإجابة.

61 بدائل المعاينة الاحتمالية

فيما يأتي بعض الطرق المعروفة للمعاينة غير الاحتمالية

1. جرى حصر العينة في جزء من المجتمع الذي يمكن الوصول إليه بسهولة. نقوم بسحب العينة من أعلى السيارة التي تحمل منتجاً زراعياً. أو سيارة مفتوحة محملة بالفحم الحجري.
2. نقوم بسحب العينة كيفما اتفق. عندما نريد أن نسحب 10 أرانب من القفص ونستخدمها في التجارب المختبرية، يقوم الباحث بإدخال يده في القفص والأرنب الذي تقع يده عليها يسحبها.
3. يقوم الباحث في المجتمعات الصغيرة وغير المتجانسة بسحب وحدات نموذجية لوحدات المجتمع وتكون قريبة من متوسط المجتمع.
4. تحتوي العينة على متطوعين فقط وهذا النوع من العينات شائع في الدراسات الطبية.

إذا توافرت الشروط الملائمة، فإنَّ أياً من الطرق أعلاه يمكن أن تعطي نتائج مفيدة، ولكنها لا يحكمها معطيات نظرية المعاينة القائمة على أساس لا معلمي، باعتبار أنه لا يوجد أي دور لعملية الاختيار العشوائية، وربما تكون الطريقة الوحيدة للتأكد من جودة أو صحة النتائج التي حصلنا عليها باستخدام إحدى هذه الطرق هو بمقارنتها بنتائج المجتمع إن وجدت، أو بنتائج حصلنا عليها عن طريقة معاينة احتمالية نفذت على نفس المجتمع في دراسات أخرى، وحتى إن كانت النتائج لصالح المعاينة غير الاحتمالية لا يوجد ما يضمن أننا سنحصل على نفس النتائج في المستقبل.

الفصل الثاني

مفاهيم أساسية للمسح بالعينات

1.2 مقدمة

لقد أصبحت المجتمعات في أيامنا هذه منشغلة بالأرقام، وهنالك رغبة للتعبير عن أي أمر من أمور حياتنا اليومية من "كتب الطبخ إلى الأفكار السياسية" باستخدام الأرقام، كما أن وسائل الإعلام اليومية من إذاعة وتلفاز وصحف ومجلات أصبحت مليئة بالمعلومات الرقمية، والحقيقة أنه لا توجد قضية يمكن طرحها وإثباتها بدون استخدام المعلومات الرقمية.

ومما لا يقبل الشك أن أفراد المجتمع أصبحوا أفضل معرفة من أي وقت مضى، وذلك بسبب استخدام المعلومات الرقمية أكثر فأكثر عن العالم الذي يحيط بهم، وهذا أمر جيد، ومن هنا ظهرت الحاجة إلى معرفة هذه الأرقام وتفسيرها، وهذا يحتاج منا إلى معلومات إحصائية معينة في الإحصاء، وعلى وجه التحديد معرفة الكيفية التي جمعت بها هذه المعلومات، والتأكد من مصداقيتها وتمثيلها إلى المجتمع الذي سحبت منه ومقدار الخطأ و... إلخ.

إن معرفة هذه الأشياء أمر حسن، ولكنه لا يكفي، إذ لا بد من معرفة كيفية تفسير هذه الأرقام أو تحليلها للوصول إلى النتائج المطلوبة، كما نرى أن العينات تمثل حجر الزاوية في كل هذه الأمور، بل هي الأساس الذي يعتمد عليه أي تحليل إحصائي أو أي استنتاجات يمكن الوصول إليها من البيانات التي وجدت غالبيتها باستخدام العينات، لذلك رأينا أن نتعرف على أساسيات

المسح بالعينات حتى يكون القارئ على دراية بهذه الأمور التي تعدّ الأساس لعملية المسح بالعينات.

22 تعريفات

1. **المجتمع** (Population): هو مجال الدراسة الذي يحوي مجموعة من العناصر التي نرغب بدراستها والحصول على بعض النتائج حولها. والمجتمع لا يعني فقط مجموعة من الأفراد ولكن قد يكون مجموعة من الحيوانات أو الحقول الزراعية أو الأشجار أو سلع معينة ينتجها مصنع أو ... إلخ.
2. **العينة** (Sample): **عبارة** عن مجموعة جزئية من المجتمع أو جزء من المجتمع.
3. **المسح الشامل** (Census survey): هو عملية عد أو حصر جميع عناصر المجتمع التي تخضع لصفات أو متغيرات معينة.
4. **وحدة المعاينة** (Sampling Unit): هي جزء مميز من المجتمع أو العنصر الذي تتم ملاحظته وإجراء القياسات وتسجيلها حوله. ووحدة المعاينة قد تكون طالباً في مدرسة أو جامعة، أو رأس غنم في قطيع أو قطعة أرض في قرية ... إلخ.
5. **تقدير** (Estimate): **هو** عبارة عن قيمة رقمية يجري حسابها من البيانات التي حصلنا عليها باستخدام العينة، وتستخدم بدلاً من معلمة المجتمع (Population Parameter) المراد تقديرها والتي تكون غير معروفة.
6. **عشوائي** (Random): تعني اختيار وحدات من المجتمع بطريقة موضوعية (تسمى طريقة المعاينة العشوائية) مثل استخدام جدول الأعداد العشوائية أو استخدام الحاسب الآلي أو أي طريقة من الطرق العشوائية الأخرى.

7. **العينة العشوائية** أو العينة الاحتمالية (Random Sample): هي عبارة عن عينة تم اختيارها **باستخدام** طريقة المعاينة العشوائية.
8. **عشوائية المعاينة** (Random Sampling): تعني بصورة عامة أي طريقة تستخدم العملية العشوائية في اختيار وحدات العينة من المجتمع.
9. **المعاينة بالإرجاع أو دون إرجاع**

(Sampling with Replacement or without Replacement): تُعنى بعملية معاملة الوحدات خلال مدة الاختيار، فإذا سحبنا الوحدة من المجتمع ولم نعدها إلى المجتمع سميت معاينة دون إرجاع، ولكن إذا تم إعادة الوحدة إلى المجتمع مرة ثانية ومن ثم قمنا بسحب وحدة جديدة تسمى معاينة مع الإرجاع.

3.2 لماذا نستخدم العينات وما مجالات استخدامها؟

تستخدم العينات بصورة واسعة في مجالات الحياة المختلفة لجمع معلومات تحتاج إليها جهات مختلفة، كالدوائر الحكومية أو المصانع أو الشركات التجارية أو الأفراد ... إلخ وفي الحقيقة كل واحد منا يستخدم العينات بصورة أو بأخرى، فمثلاً: إذا أردنا شراء سلعة معينة فإننا قبل أن نقوم بالشراء نقوم بجمع معلومات عن سعر هذه السلعة لدى بعض المحلات التي تبيعها، ومن ثم نقوم بشراء هذه السلعة، كذلك رب البيت الذي يقوم بشراء فاكهة معينة فإنه يقوم بعملية انتقاء هذه الفاكهة من بين المعروض من الفاكهة بناء على جودتها وسعرها ... إلخ، والأمثلة من هذا النوع كثيرة. ويتضح لنا من ذلك أننا استخدمنا العينات في شراء السلعة أو اختيار الفاكهة.

تستخدم العينات الآن بصورة واسعة في غالبية حقول المعرفة أو مجالات الحياة اليومية؛ ولذلك فليس من السهل أن نعطي تفصيلاً كاملاً عن جميع مجالات استخدام العينات. ولكن هنالك حالات يكون استخدام العينات فيها ضرورياً وواضح المزايا منها:

1- عندما نريد الحصول على نتائج دقيقة وذات ثقة عالية، لا سيما إذا كنا مقيدين بمبلغ محدود من المال.
2- عندما تكون الوحدات التي ندرسها ذات تشتت عالٍ بالنسبة إلى المتغيرات التي نريد دراستها.
3- عندما لا نستطيع القيام بالمسح الشامل وذلك لكونه عالي التكاليف.
4- عندما يكون مجال البحث واسعاً جداً والمجتمع غير معروف بصورة كاملة.
5- عندما نحتاج إلى نتائج سريعة لاتخاذ قرارات معينة لا يمكن الحصول عليها من خلال المسح الشامل، لأنه يحتاج إلى وقت طويل للإعداد والتنفيذ.
6- عندما تكون الموارد المالية والبشرية وكذلك الوقت غير كافية للقيام بالمسح الشامل.
7- عندما تتسبب عملية المسح في إتلاف وحدات المجتمع، ومثال ذلك معرفة نسبة البيض التالف من إنتاج مزرعة للدواجن.
8- قد يكون المجتمع غير قابل للعد مثلاً، كمخزون المملكة العربية السعودية من النفط أو مخزون الأردن من الفوسفات، ولمعرفة هذا المخزون يجب أن ننقب جميع الأراضي التابعة لهذين البلدين وهذا أمر غير ممكن.

على الرغم من المزايا الكثيرة لاستخدام العينات إلا أنها لا تخلو من عيوب أهمها:

1- مهما بلغت نتائج العينة من الدقة تبقى تقديرية وهي ليست النتائج الحقيقية للمجتمع.
2- المسح بالعينات لا يغطي إلا مجموعة جزئية من المجتمع وليس المجتمع بأكمله.

3- إن استخدام العينات يحتاج إلى أشخاص مدربين ومؤهلين علمياً للقيام بإعداد وتصميم العينات، وبدونهم لا يمكن الاعتماد على نتائج العينة.
4- تخطيط وتنفيذ المسح بالعينات يجب أن يتم بشكل دقيق جداً وإلا ستكون النتائج التي حصلنا عليها مضللة.

4.2 المسح الشامل والمسح بالعينة

إذا قمنا بدراسة جميع وحدات المجتمع بالنسبة لمتغيرات معينة فإن هذا ما يعرف بالمسح الشامل، ومما لا شك فيه أن المسح الشامل يحتاج إلى وقت طويل، وجهد كبير، وتكاليف عالية لتنفيذه، لذلك في كثير من الأحيان يكاد المسح الشامل يكون غير ممكن، وهنالك حالات (حتى ولو توفرت الجهود والتكاليف والوقت الكافي) لا يمكن معها القيام بالمسح الشامل، لعدم إمكانية تطبيقه، كما هو الحال في مثال معرفة مقدار مخزون النفط في السعودية، أو مخزون الفوسفات في الأردن.

فإذا قمنا باختيار مجموعة جزئية من المجتمع ومن ثم نقوم بجمع البيانات حول هذه المجموعة فإن هذا ما يسمى بطريقة المسح بالعينة، ومن الواضح أن المسح بطريقة العينة يحتاج إلى وقت وجهد ومال أقل بكثير من المسح الشامل. وهذا لا يعني بالضرورة أن السبب الاقتصادي هو الذي يقف خلف عملية اختيار المسح بالعينة، ولكن الأهم من ذلك لا بد من التأكد من دقة المعلومات التي حصلنا عليها سواء كان ذلك عن طريقة المسح الشامل أو المسح بالعينة. في بعض الأحيان تكون طريقة المسح بالعينة أكثر دقة من المسح الشامل. كذلك نستطيع التأكد من دقة المعلومات التي حصلنا عليها باستخدام المسح بالعينة. ولكن هذا غير ممكن بالنسبة إلى المسح الشامل حيث لا توجد طريقة للتأكد من صحة أو دقة المعلومات التي حصلنا عليها عن

طريق المسح الشامل، ومن الأمور الثابتة أن طريقة المسح بالعينة أفضل من المسح الشامل من حيث دقة النتائج التي نحصل عليها شريطة أن تكون العينة نفذت بشكل دقيق ومتقن، ولمزيد من المعلومات حول هذه النقطة يراجع (1963) Lahiri و (1953) Yates، (1961) Zarkovich و Mahalanobis (1950).

لقد لخص (1977) Cochran ميزات المسح بالعينة على المسح الشامل بأنها: تقليل تكاليف المسح، وسرعة الحصول على النتائج، ودقة أكبر في النتائج ومدى أوسع، وكذلك سهولة تعديل وتبديل المسح بالعينة.

وكذلك لخص أبو الإحصاء (1953) Fisher ميزات المسح بالعينة على المسح الشامل قائلاً: كنت أدعي أن هنالك أربعة أسباب لتفضيل المسح بالعينة. أما الثلاثة الأولى فهي سهولة التعديل والتبديل وسرعة التنفيذ وقلة التكاليف، أما الرابع فلا أحتاج لإضافة أي سبب آخر.

ولا بد من التأكيد هنا على أنّ المزايا التي ذكرناها لتفضيل المسح بالعينة على المسح الشامل لا يمكن أن تؤتي ثمارها إلا:

1. أن تكون وحدات المعاينة سحبت بطريقة صحيحة من المجتمع.
2. أن تكون طريقة المعاينة التي تم استخدامها ملائمة للمجتمع.
3. أن يكون حجم العينة الذي تم اختياره من المجتمع كافياً.

إذا كانت المعلومات المطلوب جمعها تخص جميع عناصر المجتمع فلا بد من استخدام طريقة المسح الشامل، كما هو الحال في التعدادات السكانية التي تقوم بها الدول كل عشر سنوات، بناء على توصيات الأمم المتحدة.

5.2 الخطوات الأساسية للمسح بالعينة

1- تحديد الأهداف

إن الخطوة الأولى التي يجب اتخاذها قبل القيام بالمسح بالعينة هي تحديد الهدف أو الأهداف من المسح، ولا بد من التأكد من أن هذه الأهداف متلائمة مع ما هو متوافر للمسح من أموال وأفراد ووقت.

2- تعريف المجتمع

لا بد أن يُعرف المجتمع الذي سنقوم بدراسته بطريقة واضحة لا تقبل التأويل أو عدم الوضوح، على سبيل المثال إذا أردنا تقدير مصروفات العائلة الشهرية لمدينة مكة المكرمة لا بد من تحديد وبشكل واضح حدود مدينة مكة المكرمة كأن نأخذ مثلاً الحدود البلدية المعتمدة، أو أي طريقة أخرى يمكن من خلالها تحديد حدود المدينة، كذلك لا بد أن يكون المجتمع الذي تسحب منه العينة وهو الذي يسمى **مجتمع المعاينة** مماثلاً للمجتمع الذي نهدف لدراسته وهو الذي يسمى **مجتمع الهدف**؛ ولذلك لا بد من تحديد الحدود السكانية والجغرافية والإدارية لهذا المجتمع.

3- تحديد إطار المعاينة ووحدات المعاينة

من المتطلبات الأساسية للمسح بالعينة تحديد إطار المعاينة، وهو عبارة عن كشف لجميع وحدات المجتمع، كذلك الإشارة إلى الوقت المناسب لجمع البيانات من هذه الوحدات، وتجدر الإشارة هنا إلى أن الإطار هو الذي يؤدي دوراً بارزاً في تحديد كيفية تخطيط وتصميم وبناء المسح بالعينة. كما يُعد الأساس الذي تدور حوله عملية اختيار الوحدات من المجتمع، وكذلك تقدير معلمات المجتمع. كذلك لا بد أن يكون المجتمع، قابلاً للتقسيم إلى وحدات

تسمى وحدات المعاينة، وهي وحدات متميزة عن بعضها وواضحة وغير متداخلة، وتغطي جميع المجتمع.

4- اختيار إحدى طرق المعاينة

إن اختيار إحدى طرق المعاينة المناسبة يؤدي إلى نتائج نهائية يمكن الاعتماد عليها، لذلك لا بد من إعطاء عناية خاصة لاختيار حجم العينة، وطريقة اختيار الوحدات من المجتمع وتقدير معلمات المجتمع التي هي من الأمور الإحصائية الأساسية التي يجب أن تعطى عناية كافية، كذلك لا بد من الأخذ بعين الاعتبار دقة النتائج والتكاليف عند اختيار طريقة المعاينة المناسبة.

5- تنظيم العمل الميداني

إن الوصول إلى النتائج التي نرجو تحقيقها عن طريق استخدام المسح بالعينة يعتمد إلى حد كبير على دقة تنفيذ وجمع البيانات في الميدان، وعليه فلا بد من إنجاز العمل في الميدان بصورة دقيقة وصحيحة ومخلصة؛ وبناء على التعليمات التي وضعت من قبل مصممي المسح، لذلك لا بد من تدريب العاملين في ميدان جمع البيانات ومراقبتهم في أثناء قيامهم بالعمل، وكذلك مراقبة وتفتيش البيانات التي جمعوها بدقة وبصورة متواصلة في أثناء عملية جمع البيانات وبعد الحصول عليها.

6- تلخيص وتحليل البيانات

إن الخطوة الأخيرة في عملية المسح بالعينة هي تحليل البيانات والحصول على نتائج من العينة ومن ثم تعميمها على المجتمع، وتعد هذه الخطوة أساسية وحيوية. وبما أن نتائج المسح تعد الأساس في عملية وضع السياسات – وهذا

يعتمد بصورة مباشرة على النتائج والتحليلات التي يتم التوصل إليها من البيانات – لذلك لا بد من إعطاء هذا الجانب أهمية كبيرة.

ويمكن تلخيص هذه الفقرة فيما يأتي:

أ. المحافظة على سرية البيانات.

ب. مراجعة البيانات وحذف الأخطاء الواضحة فيها وإدخالها إلى الحاسب الآلي.

ت. وضع هذه البيانات في جداول.

ث. القيام بمعالجة هذه البيانات بإجراء التحليلات الإحصائية المناسبة.

ج. كتابة تقرير في النتائج والاستنتاجات النهائية.

6.2 المعاينة الاحتمالية والمعاينة الاعتباطية (غير الاحتمالية)

المعاينة الاحتمالية هي الطريقة التي نقوم بها باختيار العينات وفقاً لبعض قوانين الاحتمالات، بحيث إن كل وحدة من وحدات المجتمع يكون لها احتمال محدد بالظهور في العينة. كذلك كل عينة من العينات الممكنة يكون لها احتمال محدد بالظهور، ومما تجدر الإشارة إليه هنا أن غالبية العينات المستخدمة هنا يكون فيها احتمال ظهور أي عنصر من عناصر المجتمع في العينة متساوياً، وكذلك احتمال اختيار عينة من بين العينات الممكنة الظهور متساوية، كما أننا سوف نقوم فقط بدراسة العينات الاحتمالية.

المعاينة الاعتباطية (غير الاحتمالية) هي طريقة لاختيار عينة أو عينات من المجتمع، بحيث إن عملية اختيار وحدات العينة يعتمد بصورة مباشرة على الشخص الذي يقوم بعملية الاختيار أو السحب، إن هذه الطريقة تُعد غير موضوعية أو شخصية، في هذه الطريقة يقوم الشخص المعاين بفحص جميع

عناصر المجتمع ومن ثم يقوم باختيار وحدات من المجتمع تكون قريبة من متوسط أو وسط المجتمع، وتستخدم هذه الطريقة للقيام بمعرفة آراء الآخرين في قضية معينة، ولكن لا يمكن بأي حال من الأحوال أن يوصى بهذه الطريقة؛ لأنها تعتمد بالدرجة الأولى على الشخص الذي يقوم بالمعاينة، وأكثر الناس لا يطمئنون لهذا الأسلوب؛ لأنه لا يخلو من التحيز من قبل المعاين، ومع ذلك إذا كان المعاين ذا خبرة ودراية وموضوعية فإنه يمكن الاستفادة من هذا النوع من المعاينة للحصول على نتائج مفيدة، ومع هذا فإنها تعاني من عيب كبير ألا وهو عدم القدرة على قياس درجة دقة التقديرات التي حصلنا عليها باستخدام هذا الأسلوب من المعاينة.

7.2 وحدة المعاينة

لقد سبق أن عرفنا وحدة المعاينة لكنها تحتاج إلى مناقشة بصورة أعمق؛ لأنها تشكل الأساس في عملية المعاينة، فالوحدة قد تكون وحدة طبيعية للمجتمع كالفرد في المنطقة السكنية، وقد تكون بتجمع طبيعي لهذه الوحدات كالعائلة، وقد تكون وحدة مصطنعة كالحقل أو قطعة الأرض ... إلخ. وقبل أن تبدأ بسحب العينة، يجب أن يكون المجتمع مقسماً إلى أجزاء متميزة وواضحة وغير متداخلة، ويجب أن تحدد وحدات المعاينة بحيث إن كل عنصر من عناصر المجتمع يقع في وحدة واحدة فقط من وحدات المعاينة، وخلاف ذلك تكون بعض العناصر لا تقع في أي وحدة ممكن سحبها من المجتمع، فعلى سبيل المثال وحدة المعاينة العائلة يجب أن تحدد بحيث إن أي شخص لا يمكن أن يكون في أكثر من عائلة، ولا يترك أي شخص دون عائلة.

8.2 الإطار أو إطار المعاينة

سبق أن عرفنا الإطار بأنه عبارة عن كشف كامل لجميع وحدات المعاينة التي تمثل المجتمع قيد الدرس، وتعد عملية بناء إطار المعاينة من الأمور الصعبة التنفيذ من الناحية الواقعية، وبصورة عامة يعد الإطار مثالياً إذا كان كاملاً وشاملاً ويحتوي على أحدث التطورات الحاصلة في المجتمع بالنسبة إلى جميع وحدات المعاينة؛ لذلك لا بد أن تجري عملية تحديث الإطار باستمرار، وأن يكون الإطار خالياً من الأخطاء ولا يهمل أي وحدة من وحدات المجتمع، ولا يكرر بعض الوحدات أكثر من مرة واحدة، وهناك نواحٍ مختلفة بالنسبة إلى إطار المعاينة جرى مناقشتها من قبل Yates (1960) و Seal (1962) و Singh (1978) و Mahalanobis (1944) وغيرهم، على أن هنالك بعض العيوب التي يمكن أن نلاحظها في إطار المعاينة منها:

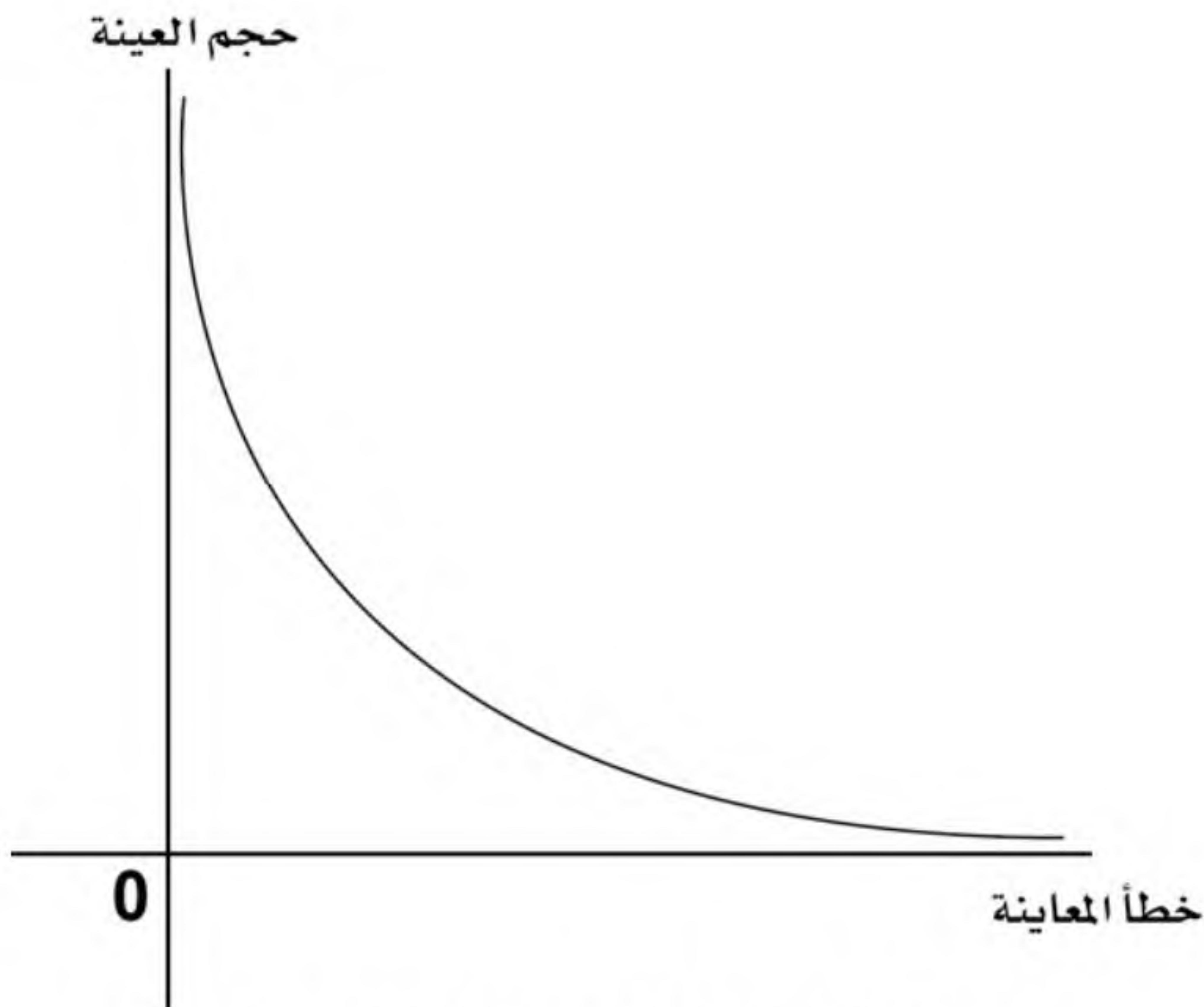
1. قد يكون الإطار غير كامل وذلك عندما تكون بعض وحدات المعاينة قد أُهملت بصورة كاملة أو كررت أكثر من مرة.
2. قد يكون الإطار غير دقيق، وذلك بأن تكون بعض الوحدات قد سجلت بصورة غير صحيحة أو أن بعض الوحدات غير الموجودة سجلت في الإطار.
3. وقد يكون الإطار غير كافٍ عندما لا يحتوي على جميع طبقات المجتمع التي يجب أن تدخل في عملية المسح.
4. قد يكون الإطار قديماً ولم تجر عملية تحديثه على الرغم من أنه كان كاملاً ودقيقاً وكافياً في الوقت الذي أُعد فيه.

ولا بد من الإشارة هنا إلى إن الإطار صعب التنفيذ بصورة دقيقة وصحيحة، ولكن الخبرة والمثابرة دائماً تؤديان دوراً في عملية بناء الإطار المتكامل.

9.2 أخطاء المعاينة والأخطاء الأخرى

1- أخطاء المعاينة

إن الخطأ الذي يحدث بسبب استخدام العينة أو البيانات التي حصلنا عليها من العينة لتقدير معلمات المجتمع يسمى خطأ المعاينة، فمهما كانت الطريقة التي نختار فيها العينة سيبقى هنالك فرق بين معلمات المجتمع (Population Parameters) وما يقابلها من تقدير لهذه المعلمات باستخدام العينة. إن خطأ المعاينة موروثٌ ولا بد منه، كما أن العينة التي يكون مقدار الخطأ فيها قليلاً تعد ممثلة جيدة للمجتمع الذي سحبت منه، ويمكن باستمرار تقليص خطأ المعاينة وذلك بزيادة حجم العينة، في الحقيقة إن حجم العينة يتناسب عكسياً مع الخطأ كما هو موضح في الشكل 1.1 ويلاحظ أنه عندما يكون حجم العينة مساوياً لحجم المجتمع يكون الخطأ صغيراً جداً أو صفراً.



الشكل 1 – 1: العلاقة بين خطأ المعاينة وحجم العينة

2- الأخطاء الأخرى

بالإضافة إلى خطأ المعاينة هنالك أخطاء أخرى يمكن إجمالها فيما يأتي:

1. عدم قياس بعض وحدات العينة التي سحبت.
2. أخطاء في الملاحظة قد يكون سببها عدم الدقة في القياس أو عطل بعض أجهزة القياس المستخدمة.
3. أخطاء بشرية مثل خطأ في تدوين بعض الأرقام أو جدولة بعض النتائج بصورة خاطئة.

على الأرجح أن تزداد هذه الأخطاء مع زيادة حجم العينة وهذا على العكس من خطأ المعاينة، بصفة عامة المسح الشامل أكثر عرضة إلى أخطاء من هذا النوع، وفي المقابل لا يوجد خطأ للمعاينة بالنسبة للمسح الشامل.

10.2 جداول الأعداد العشوائية

يُعد مفهوم العشوائية من الأمور المهمة جداً للمعاينة، وخصوصاً إذا أردنا استخدام العينة للوصول إلى نتائج، ومن ثم تعميمها على المجتمع، والسبب في ذلك يعود إلى أننا نريد أن نطبق قوانين الاحتمالات للقيام بالاستنتاج الإحصائي، إن العشوائية لا تعني التخبط في الاختيار، ولكنها تعني طريقة معرفة جيدة وبدقة للقيام بعملية الاختيار لوحدات العينة من المجتمع. والعشوائية نعني بها هنا هي الطريقة المنتظمة التي تعطي كل وحدة في المجتمع احتمالاً معلوماً أو فرصة معلومة في الاختيار من المجتمع، في أكثر طرق المعاينة العشوائية يكون احتمال جميع الوحدات بالاختيار متساوياً، ولكن لاشك أن هناك بعض طرق المعاينة يكون احتمال اختيار الوحدات فيها غير متساو، كما هو الحال بالنسبة إلى طرق المعاينة في الفصل العاشر من هذا الكتاب.

توجد طرق كثيرة لاختيار الوحدات عشوائياً من المجتمع، ومن الطرق الشائعة الاستخدام هي استخدام جداول الأعداد العشوائية، وهذه الجداول معروفة ومتداولة راجع ملحق رقم (3)، وكذلك الطريقة الأخرى التي أصبحت الآن أكثر استخداماً بسبب انتشار الحاسب الآلي حيث توجد حقائب إحصائية جاهزة يمكن استخدامها في عملية السحب العشوائية مثل MINTAB راجع الملحق رقم (2). على أنه ليس من الصعب كتابة برنامج للحاسب ليقوم بعملية الاختيار العشوائي للوحدات من المجتمع.

إن جداول الأعداد العشوائية المطبوعة عبارة عن مجموعة مكونة من الأرقام 0, 1, 2, □ , 9 اختيرت عشوائياً باستخدام إحدى طرق العشوائية المعروفة، ووضعت على شكل جدول مكونة من صفوف وأعمدة. تحتوي الجداول على نفس العدد من كل رقم من دون أي دورة أو ترتيب معين للأرقام. إن الرقم الموجود في موضع معين حدد بصورة مستقلة عن الأرقام في المواضع الأخرى، وفي العادة ترقم الصفوف والأعمدة في كل صفحة ليسهل استخدامها في جمع الأعمدة المتجاورة مع بعض إذا كانت هناك حاجة لتكون عدداً مؤلفاً في عدة أرقام.

قبل بدء عملية سحب العينة لا بد من تحديد الحد الأعلى لحجم المجتمع؛ وذلك ليكون لدينا عدد كافٍ من الأرقام في كل عدد وذلك لإعطاء جميع الوحدات الفرصة نفسها في الظهور، على سبيل المثال إذا كانت المجموعة التي نسحب منها العينة بين 100 و1000 وحدة. لذلك لا بد من استعمال 3أرقام في العدد الواحد. وإذا كانت المجموعة بين 1000 و10.000 وحدة لا بد من استخدام 4 أرقام في العدد الواحد وإذا كان حجم المجموعة غير معروف نحتاج إلى الحد الأعلى للمجموعة.

لاستخدام جداول الأعداد العشوائية نبدأ في نقطة تسمى نقطة البداية. كثير من الكتاب يقترحون فتح صفحة من جداول الأعداد العشوائية، ومن ثم نقوم برمي القلم داخل الجدول والرقم الذي يقع عليه القلم يكون هو نقطة البداية، يبدو أن هذه الطريقة غير علمية، وقد تؤدي إلى أن جميع المسوحات تبدأ بالمنطقة نفسها في الجدول، كطريقة بديلة يمكن أن تحدد نقطة البداية بسحب أحد الأعمدة وأحد الصفوف بصورة عشوائية، والرقم الذي يلتقي عنده العمود والصف المختاران بصورة عشوائية يمثل نقطة البداية. والخطوة الأخيرة هي تحديد العلاقة بين العدد العشوائي الذي اختير من الجدول وعناصر المجتمع. وهذا يكون إذا سحب الرقم 4256 مثلاً فإنه يعني أن وحدة معينة من وحدات المجتمع اختيرت ونستطيع الآن تحديدها والوصول إليها ومن ثم جمع المعلومات منها.

مثال: لنفترض أننا نريد أن نسحب عينة عشوائية متكونة من 5 وحدات من مجتمع يحتوي على 200 وحدة، ولتكن هذه الوحدات تمثل عائلات واقعة في منطقة سكنية معينة، تبدأ عملية سحب عمود وصف بطريقة عشوائية وحيث إن في الجدول الموجود في الملحق رقم 1 توجد في الصفحة 13 أعمدة كل عمود يحتوي على 4 أرقام أي في الحقيقة لدينا 52 عموداً و 51 صفاً في الصفحة، نقوم باختيار عمود من بين الأعمدة وصف من بين الصفوف بطريقة عشوائية. لنفترض أنه ظهر لدينا العمود رقم 27 والصف رقم 34 لذلك يكون الرقم 3 هو نقطة البداية، ولقد رُقِّمت وحدات المجتمع 001، 002، 003، □ 200 وعليه سوف تختار كل مرة عدداً مكوناً من 3 أرقام، كما ذكرنا فإن نقطة البداية هي الرقم 3 الواقع عند تلاقي العمود رقم 27 والصف رقم 34. الآن نبدأ التحرك نحو إما يمين أو يسار، كل مرة نأخذ 3 أرقام 360، 260، 406، وهكذا. سوف نهمل الأعداد التي تكون أكبر من 200، ونأخذ الأعداد التي هي أقل أو تساوي 200؛ لأن مجتمعنا يتكون من 200 وحدة. لذلك ستكون

الوحدات رقم 166، 124، 068، 060، 092 في العينة. طبعاً إذا انتهى صف ننتقل إلى الصف الذي يليه وهكذا. إذا انتهت الصفحة نرجع إلى بداية الصفحة وهكذا. من الملاحظ أن كثيراً من الأعداد أُهملت لأنها غير داخلة في المجتمع. يمكن استخدام بعض الطرق الرياضية للتقليل من عدد الأعداد التي تهمل، وبالإمكان سحب هذه العينة بسهولة جداً وذلك باستخدام الحقيبة الإحصائية MINTAB انظر الملحق رقم 2.

المراجع العربية

1. وليم كوكوران – ترجمة أنيس كنجو – تقنية المعاينة الإحصائية (1995)، جامعة الملك سعود، الرياض.
2. حسين علوان (1994)، طرق المعاينة، دار الفرقان، عمان.

References

1. Barnett, V. (1991). Sampling Survey Principles and Methods, Edward Arnold, London.
2. Cochran, W. G. (1977). Sampling Technique, 3rd Ed. John Wiley & Sons, New York.
3. Fisher, R. A. (1953). The Design of Experiments, Oliver & Boyd, London.
4. Lahiri, D. B. (1951). A Method of Sample Selection Providing Unbiased Ratio Estimates Bull. Int. Statist., 33, 133, 140.
5. Levy, P. S. and Lemeshow, S (1999). Sampling of Population Methods and Applications, Wiley, New York.
6. Mahalanobis, P. C. (1944). On the Large Scale Sample Surveys, Phil. Trans. R. Soc., 231, 329 451.
7. Mahalanobis, P. C. (1950). Cost and Accuracy of Results in Sampling and Complete Enumeration, Bull. Int. Statist. Inst., 32, 210 213.
8. Scheaffer, R. L., Mendenhall, W. and Ott, L. (1996). Elementary Survey Sampling, 5th Ed., Duxbury press, Boston.
9. Seal, K. C. (1962). Use of Outdated Frames in Large Sample Surveys. Bull Cal. Statist Assoc., 11, 68 84.
10. Singh, D. (1978). Taking Agricuitural Censuses, FAO, Rome.
11. Sturat, A. (1984). the Ideas of Sampling, Revised Edition, Griffin, London.
12. Thompson, S. K (2002). Sampling, 2end Wiley, New York.
13. Yates, F. and Grundy, P. M. (1953). Selection without Replacement from Within Strata Probability Proportional to Size J. R. Statist, Soc. 15 B, 253 261.
14. Yates, F. (1960). Sampling Methods for Censuses and Surveys, Charles Grittin & Co., London.
15. Yates, F. (1981). Sampling Methods for Censuses and Surveys 4th Ed., Grittin, London.
16. Zarkovich, S. S. (1961). Sampling Methods and Censuses, FAO, Rome.

الفصل الثالث

متطلبات العينة الجيدة

1.3 مقدمة

إن الهدف الرئيس لأي طريقة من طرق المعاينة هو الحصول على عينة يمكن من خلالها الوصول إلى صفات المجتمع تحت الدراسة وبخاصة تلك التي تهمنا بالدرجة الأولى، أو التي يهدف المسح بالعينة الحصول على معلومات عنها. فمثلاً إذا أردنا الحصول على معلومات عن مصروفات العائلة الشهرية على المواد الغذائية في بلد معين أو منطقة معينة، فإن المعلومات التي تهمنا بالدرجة الأولى هي: دخل العائلة الشهري، أو معدل دخل العائلة الشهري، ومعدل مصروف العائلة على المواد الغذائية، ولكن قد تكون هنالك صفات أخرى للمجتمع نريد معلومات عنها مثل معدل عدد أفراد العائلة الواحدة، ومعدل عدد العاملين في العائلة، ومعدل عدد الأطفال دون سن 18 سنة ... إلخ. ولكن ليست هي الهدف الرئيس للمعاينة، ولكن هذا لا يعني أن نصل إلى صفات المجتمع بأي وسيلة وإنما ضمن الشروط والحدود المفروضة أو الموضوعة للبحث، على سبيل المثال قد نحدد وقتاً نهائياً لإنجاز المسح، أو قد تكون ميزانية الدراسة محدودة ... إلخ.

23 التحيز

قد يخطر على بال كثير منا أنه إذا أردنا تقدير معدل إنتاج الدونم الواحد من القمح أن نسأل أحد ذوي الخبرة أو أحد الموظفين العاملين في دائرة الزراعة

كـي يخـرج إلى حقـول القمـح ويختـار بعـض الحقـول ذات الإنتاجيـة المتوسـطة، ويحـدد إنتاجيـة هـذه الحقـول، ومـن ثـم نسـتطيع أن نحـدد معـدل إنتـاج الـدونم الواحـد مـن القمـح في منطقـة معينـة بهـذه الطريقـة. تبـدو هـذه الطريقـة للوهلـة الأولى أنها جيدة لأنها تختار حقولاً متوسطة الإنتاجية، ونحن هدفنا تحديد معدل إنتاج الدونم الواحد. إن هذه العينة أو هـذه النوعيـة مـن المعاينـة لا يمكـن بأي حال من الأحوال الاعتماد عليها، لأنها لا قيمة لها. ولكن خطؤها الوحيد أنهـا يمكـن أن تكـون متحيـزة، وهـذا يعـني أن عمليـة اختيـار الحقـول ذات الإنتاجية المتوسطة معرضة للخطأ في الاختيار، وذلك لعدة أسباب منها:

1. قـد يكـون الخبير أو موظـف دائـرة الزراعـة في المنطقـة يميـل إلى اختيـار الوحدات ذات الإنتاجية العالية، لأنـه يرغـب أن تظهـر منطقتـه علـى أنهـا أفضل المناطق في الإنتاج، أو أنها سياسة عامة من قبل الدولة للميل نحو إظهـار أن إنتاجيـة البلـد منخفضـة لأسـباب سياسـية أو أمنيـة أو ... إلخ. وكـذلك فقـد يشـعر الخبير أو الموظـف بأنـه لـيس مـن مصـلحة الفلاحـين إعطاء إنتاجية عالية لحقولهم، لأسباب قد تكون ضريبية أو لمساعدات تقدمها الدولة إلى الفلاحين.
2. حتى ولو سلمنا أن الخبير أو الموظف سوف يقوم بواجبه ويعطي الأرقام الحقيقـة مهمـا كانـت الظـروف، تبقـى هنالـك أخطـاء غـير ممكنـة السـيطرة عليهـا، كالأخطـاء الشخصـية لأنـه تبقـى العمليـة اختيـارًا شخصيًا. وممكن أن يقع الشخص في الخطأ من دون قصد. وقد يكون خطأ تحديد حقل واحد من الحقول يلغي كل الدقة التي حصل عليها من خلال تحديد الحقول الأخرى.
3. إذا كـان الخبير أو الموظف ذا خبرة وموضوعية ويمكن الحصول على نتائج جيدة وخالية من الأخطاء حتى الشخصية منهـا، فـإن هـذه الطريقـة

لا يمكن قياس دقة التقديرات التي حصلنا عليها من خلالها، ومن ثم لا نستطيع الاعتماد على النتائج التي حصلنا عليها.

لا بد أن نفرق بين نوعين من أخطاء المعاينة، الأول سببه التحيز في الاختيار، والثاني سببه الفرق بين وحدات المجتمع الموجودة في العينة والوحدات الأخرى الموجودة في المجتمع ولكن ليست في العينة، ويسمى الأول خطأ التحيز (أو سببه التحيز) ويسمى الثاني خطأ المعاينة العشوائية أو يسمى خطأ المعاينة إذا لا يوجد تحيز في الاختيار، على أن الخطأ بسبب التحيز سوف لا يقل مهما كبر المجتمع أو كبرت العينة، ولكن خطأ المعاينة العشوائية يقل بالمعدل إذا زاد حجم العينة.

3.3 طرق اختيار تسبب التحيز

توجد هنالك طرق عديدة يمكن أن نختار بها وحدات العينة بطريقة غير صحيحة قد تسبب تحيزاً في الاختيار، وأهم هذه الطرق يمكن إجماله بصورة عامة فيما يأتي:

1. الاختيار المتعمد لعينة ممثلة للمجتمع وهذا النوع من التحيز نوقش في الفقرة أعلاه.
2. اختيار يعتمد على بعض الصفات التي تكون لها علاقة أو تكون مرتبطة مع خواص الوحدة ذات الاهتمام، إن كثيراً من الاختيار الاعتباطي يسبب تحيزاً من هذا النوع.
3. اختيار وحدات العينة من غير وجود طريقة واضحة تستعمل في اختيار وحدات العينة من بين وحدات المجتمع عشوائياً. أي عدم وضوح الطريقة العشوائية في الاختيار. لكن الباحث يدعي أنه طبق العشوائية، وربما يسمح الباحث لرغباته بإيجاد نتائج معينة يكون لها تأثير على اختياره

للوحدات، وهذا النوع من التحيز أخطر أنواع التحيز؛ لأن وجوده غير ظاهر للعيان.

4. غالباً ما يعمد الباحثون أو العادون إلى تعويض وحدة ملائمة من المجتمع بدل الوحدة الموجودة في العينة، والتي وجدوا صعوبة ما في الوصول إليها للحصول منها على المعلومات. فيؤخذ المنزل المجاور بديلاً عن المنزل الموجود في العينة عندما لا يكون هنالك جواب من المنزل المعني. هذا يؤدي إلى أن المنازل التي ستكون في العينة هي المأهولة طوال اليوم؛ أي المنازل التي يسكن فيها عائلات مع أطفالهم.

5. الفشل في تغطية كل العينة المختارة. فإذا لم تكن هنالك زيارة ثانية إلى المنازل التي لم نحصل على جواب منها في الزيارة الأولى سيكون هنالك تحيز في الاختيار، حتى ولو لم نحاول التعويض عنها في المرة الأولى، وهذه الظاهرة تظهر جلياً في المسح الذي يُجرى بواسطة البريد، حيث إنه غالباً ما يكون الجواب غير كامل من قبل كل الذين أرسلنا لهم الاستبانات، ويرد غالباً الأشخاص الذين لهم اهتمام بموضوع البحث، وهؤلاء لا يمثلون المجتمع.

4.3 تجنب التحيز في الاختيار

من الواضح أنه إذا كان هنالك تحيز في الاختيار لا يمكن الوصول إلى نتائج موضوعية من العينة؛ إذن المهمة الأولى الأساسية لأي طريقة من طرق المعاينة هي التخلص من مصادر التحيز، إن أبسط طريقة عالمية للتخلص من التحيز أو لتجنب التحيز هي أن نسحب العينة بصورة عشوائية كاملة أو عشوائية مشروطة، تؤدي إلى زيادة دقة العينة ولا تؤدي إلى التحيز في النتائج. وفي بعض الأحيان هنالك ما يسمى بالاختيار المنظم، مثل اختيار أسماء من كشف، على مسافات متساوية من بداية الكشف إلى نهايته، أو اختيار نقاط

على خريطة باستخدام المربعات المتساوية المساحات، هذا النوع من الاختيار ربما يكون مسموحاً به.

إن عشوائية الاختيار لا تعني التخبط في الاختيار، ويمكن الحصول على العينة العشوائية باستخدام بعض الطرق العشوائية الملائمة، كاستخدام جداول الإعداد العشوائية، أو استخدام الحاسب الآلي .... إلخ. ولا بد من التأكيد هنا على أن كلمة عشوائية أو عينة عشوائية يساء استخدامها كثيراً من قبل الناس، ولهذا السبب لا بد أن تشرح بصورة واضحة طريقة سحب العينة العشوائية من المجتمع في جميع الأعمال التي تستخدم العينة العشوائية.

ولأجل منع التحيز الناتج عن الإهمال أو التعمد "إذا كان العمل كبيراً نسبياً" يفضل أن ينفذ بصورة مركزية من قبل دائرة واحدة بطريقة لا تترك أي خيارات أمام العادين أو الباحثين. وكذلك أن ينفذ بطريقة يسهل معها مراقبة العمل الميداني كلما دعت الحاجة إلى ذلك.

5.3 تحيز بسبب سوء تحديد وحدات المعانية

عندما لا تكون وحدات المعاينة وحدات طبيعية للمجتمع، يجري في الغالب تحديد حدود الوحدات التي اختيرت في الوقت الذي تجري فيه عملية القياسات، إن عملية تقدير إنتاج المحاصيل الحقلية تجري عادة عن طريق اختيار وحدات صغيرة من المساحة للحصول منها على تقدير للمحصول أو أي صفات أخرى للمحصول. وتجري عملية تحديد مواقع هذه المساحات الصغيرة نظرياً من خلال اختيار نقاط عشوائية في نظام الإحداثيات، وهذا يضمن العشوائية في الاختيار، وبما أنه واقعي فمن المستحيل تحديد موضع هذه المساحات في الحقل بصورة دقيقة؛ لذلك لا بد من استخدام طرق تقريبية. ويجب أن تكون المساحات غير صغيرة في هذا النوع من العمل؛ لأن الخطأ في تحديد حدود المساحات يزداد كلما صغرت المساحة، وكذلك فإن أي تغيير

بسيط في موضع الوحدة سوف يؤثر في النتائج إذا كانت المساحات صغيرة. إن الوحدات الصغيرة تعطي نتائج دقيقة جداً إذا كان هناك عاملون ذوي خبرة في العمل الميداني، ولكن ربما تكون النتائج غير دقيقة إذا كان العاملون غير مدربين ولا توجد لديهم خبرة ميدانية.

6.3 تحيز في التقدير

بالإضافة إلى التحيز الذي ينتج عن سوء طريقة الاختيار أو تحيز سببه عملية جمع المعلومات، يوجد هنالك تحيز آخر ناتج من الأسلوب المتبع لتقدير النتائج، ومثال بسيط على هذا النوع من التحيز يحدث عند تقدير النسبة. فعلى سبيل المثال، توجد ثلاثة أنواع من الأراضي المختلفة من حيث الخصوبة، ومن ثم مختلفة من حيث معدل إنتاجية الدونم الواحد من محصول معين، فإذا كان معدل إنتاج هذه الأنواع الثلاثة من الأرض من محصول معين على التوالي 5 طن، 3 طن، 2 طن ومعدل مساحة الحقول الثلاثة على التوالي هي 5 دونم، 10 دونم و15 دونم وعدد الحقول مساوٍ للأنواع الثلاثة. إن معدل إنتاج الدونم الواحد يمكن الحصول عليه باستخدام الوسط الحسابي المرجح $\bar{Y}_w$ طن للدونم الواحد:

$$\bar{Y}_w = \frac{5(5)+3(10)+2(15)}{5+10+15}$$

بينما الوسط الحسابي لإنتاج للدونم الواحد لكل الحقول $\bar{Y}$ طن للدونم الواحد:

$$\bar{Y} = \frac{5+3+2}{3} = 3.33$$

والنتيجة وجود تحيز في التقدير من قبل الأخير$\bar{Y}$ مقداره حوالي 18%، على أن التحيز بسبب التقدير مقدور عليه من خلال استخدام طرق التقدير الملائمة.

7.3 الحالات التي يسمح بها التحيز

على الرغم من أن تحاشي التحيز يُعَدُّ موضوعاً مهماً وخصوصاً في المسوحات التي يترتب عليها قرارات إدارية معينة، إلا أنه يمكن قبول كمية معينة ثابتة من التحيز في بعض المسوحات حقب. ويعد التحيز في المسوحات التي تجري على حقب منتظمة –والتي يكون الهدف منها تحديد التغير الحاصل من حقبة إلى أخرى وليس القيمة المطلقه– مقبولاً إذا كان صغيراً وثابتاً مع الوقت. كذلك في المسوحات التي يكون هدفها الرئيس المقارنة بين مجموعات مختلفة من المجتمع، يعد التحيز الذي يكون ثابتاً من مجموعة إلى أخرى ذا أهمية قليلة، ويجب أن يتحاشى الباحثون وضع أهمية مبالغ فيها على مصادر التحيز الصغيرة، التي ينتج عنها في الحقيقة خطأ تافه قياساً إلى خطأ المعاينة العشوائية.

8.3 طرق لتقليل أخطاء المعاينة

بعد التخلص من التحيز لا بد من توجيه اهتمامنا إلى أخطاء المعاينة، هذه الأخطاء يجب أن تكون صغيرة حتى نصل إلى الدقة المطلوبة، فإن أبسط الطرق لزيادة الدقة هو زيادة حجم العينة، إذا كانت الأمور الأخرى كما هي عند زيادة حجم العينة، فإن خطأ المعاينة يتناسب عكسياً على الجذر التربيعي لحجم العينة.

إن الحصول على الدقة المطلوبة لا يعتمد فقط على حجم العينة، وإنما أيضاً على التغير لكل وحدة، بصورة أكثر تحديداً على الجزء من التغير

لكل وحدة والذي يدخل ضمن خطأ المعاينة. وباستخدام طرق اختيار ملائمة – والتي تضع قيوداً على العملية العشوائية الكاملة التي لا تسبب تحيزاً في النتائج– يمكن دائماً تقليل خطأ المعاينة بصورة كبيرة رغم بقاء حجم العينة ثابتاً، إن من أبسط أنواع القيود التي تضع على عملية الاختيار العشوائية الكاملة هي (الطبقية) التي يقسم فيها المجتمع إلى طبقات من الوحدات، بحيث تكون الوحدات في كل طبقة متماثلة، وبعد ذلك تسحب عينة عشوائية من كل طبقة. فإذا سحبنا من كل طبقة وحدات متناسبة مع حجم الطبقة سيؤدي هذا إلى إزالة الفروق بين الطبقات المختلفة، ومن ثم التقليل من خطأ المعاينة.

وبالإضافة إلى استخدام الطبقية هناك ثلاثة طرق أخرى يمكن استخدامها لزيادة الدقة، وفي بعض الأحيان تؤدي إلى زيادة كبيرة في الدقة، ومن ثم يمكن تقليل خطأ المعاينة. وهذه الطرق هي:

1. استخدام المعلومات الإضافية المتوافرة. وسوف نتكلم عن هذا الموضوع بالتفصيل في أحد الفصول القادمة.
2. استخدام العينة متعددة المراحل أو العينة العنقودية، وكذلك سوف نتكلم عن هذا النوع من المعاينة في أحد الفصول القادمة.
3. استخدام التوزيع الأمثل. إن هذه الطريقة تضمن لأجزاء مهمة أو أجزاء أكثر تغيراً من المجتمع أن تسحب منها وحدات أكثر. إن التوزيع الأمثل للعينة يعتمد على التغير النسبي للطبقات المختلفة التي تقسم المجتمع، فمثلاً إذا كان مطلوباً أن نحدد عدد العاملين في صناعة معينة فمن الأفضل أن نأخذ عدداً أكبر من المعامل الكبيرة وأقل من المعامل الصغيرة.

9.3 اختيار الوحدة

توجد اختيارات كثيرة ومتنوعة في اختيار نوع وحجم وحدات المعاينة في بعض أصناف المواد، وهذا يعطي مجالاً واسعاً لزيادة كفاءة طريقة المعاينة. وبصفة عامة كلما صغر حجم الوحدة من المواد الداخلة في العينة زادت دقة النتائج، فعلى سبيل المثال في المسح الزراعي سيكون أكثر دقة أن نأخذ 10% من جميع الحقول الموجودة في كل وحدة إدارية من أن نأخذ جميع الحقول الموجودة في 10% من الوحدات الإدارية، ويبقى هذا الكلام صحيحاً حتى في حالة استخدام العينة متعددة المراحل.

إن الحاجة إلى وحدات صغيرة موزعة على كل أجزاء المجتمع غالباً ما يتناقض مع المتطلبات الإدارية، واضح لنا أنه من السهل الإعداد لمسح حقول في مناطق صغيرة مثل الوحدات الإدارية أو المقاطعات من الإعداد لمسح بنفس العدد من الحقول موزعة على كل أجزاء البلد. إن الاختبار المتوازن بين هذين الأمرين المتناقضين غالباً ما يكون هو إحدى المشكلات الرئيسة للتخطيط للمسح بالعينة. فإذا قمنا بسحب عدد صغير من الوحدات الكبيرة في العينة فإن ذلك سيؤدي إلى عدم تحديد خطأ المعاينة بصورة جيدة؛ لأنه لا يوجد سوى فرق قليل نسبياً بين الوحدات الكبيرة.

ولا بد من الإشارة هنا إلى أن اختيار طريقة المعاينة لا يعتمد فقط على دقتها بالنسبة إلى الطرق الأخرى، ولكن يعتمد كذلك على مدى ملاءمتها للتطبيق على أرض الواقع، إن من الأهمية بمكان أن تعتمد طريقة المعاينة على ما هو متوافر في المجتمع من المعلومات والوسائل، فمثلاً قد تكون طريقة معينة ممتازة في بلد توجد فيه خرائط، بالمقابل تكون عديمة الفائدة في بلد آخر لا توجد فيه خرائط.

المراجع العربية

1. وليم كوكوران – ترجمة أنيس كنجو – تقنية المعاينة الإحصائية (1995)، جامعة الملك سعود، الرياض.
2. حسين علوان (1994)، طرق المعاينة، دار الفرقان، عمان.

References

1. Barnett, V. (1991). Sampling Survey Principles and Methods, Edward Arnold, London.
2. Cochran, W. G. (1977). Sampling Technique, 3rd Ed. John Wiley & Sons, New York.
3. Fisher, R. A. (1953). The Design of Experiments, Oliver & Boyd, London.
4. Lahiri, D. B. (1951). A Method of Sample Selection Providing Unbiased Ratio Estimates Bull. Int. Statist., 33, 133, 140.
5. Levy, P. S. and Lemeshow, S. (1999). Sampling of Population□Methods and Applications, Wiley, NewYork.
6. Mahalanobis, P. C. (1944). On the Large□Scale Sample Surveys, Phil. Trans. R. Soc., 231, 329□451.
7. Mahalanobis, P. C. (1950). Cost and Accuracy of Results in Sampling and Complete Enumeration, Bull. Int. Statist. Inst., 32, 210□213.
8. Scheaffer, R. L., Mendenhall, W. and Ott, L. (1996). Elementary Survey Sampling, 5th Ed., Duxbury press, Boston.
9. Seal, K. C. (1962). Use of Outdated Frames in Large Sample Surveys. Bull□Cal. Statist Assoc., 11, 68□84.
10. Singh, D. (1978). Taking Agricuitural Censuses, FAO, Rome.
11. Sturat, A. (1984) the Ideas of Sampling, Revised Edition, Griffin, London.
12. Thompson, S. K (2002). Sampling, 2end Wiley, New York.
13. Yates, F. and. Grundy, P. M (1953). Selection without Replacement from Within Strata Probability Proportional to Size J. R. Statist, Soc. 15 B, 253□261.
14. Yates, F. (1960). Sampling Methods for Censuses and Surveys, Charles Grittin & Co., London.
15. Yates, F. (1981). Sampling Methods for Censuses and Surveys 4th Ed., Grittin, London.
16. Zarkovich, S. S. (1961). Sampling Methods and Censuses, FAO, Rome.

الفصل الرابع

تنفيذ المسح بالعينة

1.4 مقدمة

بعد أن يكون الهدف أو الأهداف الأساسية من المسح قد حددت، ودرجة الدقة المطلوبة من البيانات كذلك حددت، وجميع الخطوات الأساسية مثل تحديد المجتمع ووحدة المعاينة وحجم العينة وطريقة المعاينة التي سوف تستخدم قد حددت أيضاً، تبدأ مرحلة التنفيذ التي قد تبدأ بتساؤلات منها:

1. كيف نطرح الأسئلة بصورة واضحة وغير مضللة؟
2. كيف نشجع الناس للإجابة عن أسئلتنا بدقة؟
3. كيف نتصل بالذين نرغب في مقابلتهم لجمع البيانات منهم؟
4. كيف نضمن أن مجموعة معينة من الناس والذين لهم مصالح معينة من هذا المسح أن تستجيب للإجابة عن الأسئلة، ولا ترفض بحيث تؤثر على النتائج وتجعلها متحيزة؟
5. كيف نقرر ما هو الصحيح والملائم أو المناسب لنسأل عنه؟
6. كيف نستطيع أن نقيم عملنا بحيث نقول: إننا قد حققنا أهدافنا الإحصائية من المسح وخصوصاً عدم التحيز ودقة المعلومات التي حصلنا عليها؟

نلاحظ أن غالبية التساؤلات أعلاه تدور حول المسح بالعينة والتي تخص معرفة آراء الناس حول قضية أو قضايا تهمهم، ولكن في الحقيقة إن عملية تنفيذ أي مسح بالعينة تسبقها تساؤلات عن كيفية التنفيذ، وقبل أن ننتقل في

دراستنا إلى التصميم والتحليل الإحصائي، رأينا أن نتفحص مدى المشكلات الواقعية التي سترافق عملية جمع البيانات وكيف نستطيع أن نتعامل معها ونتغلب عليها.

2.4 اختيار وتدريب العاديّن

سوف لا أدخل في تفصيلات كثيرة في هذا الموضوع، ولكن أحب أن أنوه إليه؛ لأن العاديّن لهم دور مهم في عملية المسح بالعينة، لذلك لابد أن يتم اختيارهم بشكل دقيق، وأن يكونوا ذوي خبرة وذكاء ولديهم القابلية للتعامل مع الناس ولديهم بعض المعلومات عن المبادئ الأولية في الحسابات. وبعد اختيارهم لابد من تدريبهم على عملية المسح وإطلاعهم على أهداف المسح، والقيام بإعدادهم بشكل جيد ليقوموا بعملهم على أفضل وجه. جرت العادة على أن تعقد دورات خاصة لتدريبهم وإعدادهم من قبل الدوائر أو الدائرة المشرفة على المسح، ولمزيد من المعلومات يراجع Raj (1972). ولابد من الإشارة في نهاية الحديث عن هذا الموضوع إلى ضرورة الإشراف ومتابعة العاديّن عند القيام بأعمالهم للتأكد من قيامهم بها وفق ما طلب منهم.

3.4 مصادر التشتت والخطأ

يمكن أن تكون نقطة البداية هنا بالسؤال الآتي: ما الخطأ الذي يمكن أن يحدث؟ ولماذا؟ لنفرض أننا نستطيع أن نتخذ قراراً فيما يخص ميزانية المسح. وقبل أن تبدأ عميلة المسح لا بد من التأكد أن البيانات بالشكل التي حددها التصميم يمكن جمعها، وما هي المصاعب التي يمكن أن تواجهنا في عملية تنفيذ المسح؟ يمكن أن نحصل على معلومات من خلال دراسة واختيار الموقع أو مسح أولي قبل بدء المسح الرئيس كي نستطيع أن نعدل على التصميم ليصبح ممكن التنفيذ، وهنالك مصادر متعددة للخطأ في المعاينة

يمكن أن نلخصها فيما يأتي: أخطاء المعاينة، وأخطاء التغطية أو عدم ضم جميع الوحدات، وأخطاء ناتجة عن عدم الإجابة(أخطاء الاستجابة)، وأخطاء المشاهدة، وسوف نتطرق إلى كل مصدر خطأ مما سبق على حده .

1. خطأ المعاينة

هذا النوع من الخطأ لا يمكن التخلص منه، وذلك لأنه نشأ من اختلاف وحدات المعاينة فيما بينها، فمثلاً إذا كنا نجمع بيانات عن دخل الأسرة نلاحظ أن هناك أسرة دخلها مرتفع جداً وأخرى منخفض، وهكذا لا نستطيع التخلص من هذا النوع من الفروقات، لذلك فإن خطأ المعاينة لا بد منه، ويمكن تقليله عن طريق زيادة حجم العينة أو تقسيم المجتمع إلى طبقات أو استخدام بعض المعلومات الإضافية.

2. أخطاء التغطية أو عدم ضم جميع الوحدات

هذا النوع من الأخطاء يحدث عندما تكون لدينا وحدة من وحدات المجتمع لا يمكن أن تظهر في العينة. مثلاً عندما نجري عملية المسح باستخدام الهاتف فقط فإن كل وحدات المجتمع التي ليس لديها تلفون سوف لا تكون لها فرصة في الظهور في العينة، أو أن نقوم بسحب عينة من مدينة ما في وقت تكون هنالك مباراة في كرة القدم في ملعب المدينة، واضح هنا أن الرياضيين والمشاهدين للمباراة سوف لن تكون لهم فرصة في الظهور في العينة، وهذا النوع من الأخطاء يحدث عندما يكون هناك سوء في المطابقة بين إطار المعاينة والمجتمع الذي نهدف إلى دراسته. وكلما زادت الوحدات التي لا يمكن أن يغطيها المسح زاد التحيز وتعرض البحث أو المسح إلى خطر عدم قبول نتائجه أو التشكيك في دقتها وتمثيلها للمجتمع.

3. أخطاء بسبب عدم الإجابة (أخطاء الاستجابة)

كما هو واضح من التسمية فإن عناصر من المجتمع التي ظهرت في العينة لم نستطع الحصول منها على المعلومات أو البيانات المطلوبة، وغالباً ما يكون هناك أكثر من متغير أو صفة يتم مشاهدتها في نفس الوحدة، كما هو الحال بالنسبة إلى الاستبانات التي تحتوي على مجموعة من الأسئلة، فعدم الاستجابة قد يتم مع مجموعة من الأسئلة حيث تترك دون جواب، كما هو الحال بالنسبة إلى الاستبانات التي ترسل بواسطة البريد، فمثلاً قد يرفض الشخص الإفصاح عن دخله. إن سبب عدم الاستجابة يحدث لأسباب مختلفة منها: قد يكون لطبيعة المعلومات المطلوبة حقائق أو (رأي شخصي)، كذلك قد يكون السبب هو الطريقة أو حتى الوقت الذي نطلب فيه جمع المعلومات، أو قد تكون هذه العوامل أو الأسباب متداخلة أو مترابطة مع بعضها.

غالباً ما يعد عدم الاستجابة الكلية مقياساً على نجاح أو عدم نجاح المسح، ربما يكون عدم الاستجابة سببها هو القرارات غير الصائبة عن كيف ومتى وأين وماذا نجمع، وما هي المعلومات المطلوبة. لوحظ أن موضوعات الاستبانات المختلفة والطرق المختلفة لجمع البيانات تسبب مستويات مختلفة من عدم الاستجابة، إن فقدان جزء من العينة بسبب عدم الاستجابة سيؤدي إلى تضخم التباين للتقديرات المختلفة، ولكن درجة تأثيرها على التحيز يعتمد على نوع المجموعة التي لم تستجب هل هي تمثل جزءاً معيناً من المجتمع أم لا؟. فإذا كانت تمثل رأي جزء من المجتمع فإن مقدار التحيز سوف يتضخم، إذا كان المتغير أو الصفة التي نجمع عنها المعلومات لها علاقة قوية مع عدم الاستجابة (مثلاً الذين يتقاضون رواتب كبيرة يكونون أكثر تحفظاً في الإجابة عن أسئلة تخص مدخولاتهم) نتوقع أن يكون التحيز كبيراً وخطيراً.

إن من أصعب أنواع عدم الاستجابة التي يصعب التعامل معها هو الشخصي منها والذي سببه رفض التعاون مع العادين. وهذا النوع يتأثر بصورة خاصة بالطريقة التي تستخدم (مثلاً مقابلة شخصية، استعمال البريد ...إلخ) لجمع المعلومات. إن الفشل في إيجاد عنصر من عناصر العينة يعد أحد مصادر عدم الاستجابة، ويمكن علاج هذا النوع من عدم الاستجابة بمحاولة إيجاد هذا العنصر في محاولة لاحقة.

4. أخطاء المشاهدة

يمكن أن نعرف خطأ المشاهدة: بأنه الحصول على معلومات خاطئة من وحدات المعاينة كما يمكن أن تحدث أخطاء المشاهدة من خلال طرق مختلفة:

1. خطأ المقابلة أو خطأ السؤال: عندما يوجه السؤال بطريقة خاطئة أو بطريقة غير واضحة تؤدي إلى أن تكون الإجابة خاطئة.
2. خطأ التسجيل: تكون الإجابة صحيحة ولكنها تسجل بطريقة خاطئة.
3. خطأ النقل أو الإدخال: عندما تدخل المعلومات في الحساب الآلي أو قاعدة البيانات يحدث خطأ ما في أثناء ذلك.

هذه الأخطاء ليس لها علاقة بوحدة المعاينة أو عنصر العينة وقد تكون الإجابة خاطئة عن سؤال معين حتى ولو كان هذا السؤال قد طرح بشكل واضح ودقيق، وأسباب ذلك تعود إلى أنّ السؤال له علاقة بموضوع حساس، أو له تأثير نفسي على الشخص، أو لأنه احتوى على تفاصيل كثيرة، أو لأنه قد يجرم أو يلقي التهمة على الشخص.

في نهاية المطاف لا بد من أن نعرج قليلاً على أخطاء سببها القياسات، وذلك بأن تكون أجهزة القياس غير صالحة أو لا تقرأ القياسات بشكل

صحيح، أو لا يحسن الباحث استخدام الأجهزة، هذه أمور مقدور عليها. ولكن تبقى هناك أخطاء عندما توجد قيمة معينة يصعب الحصول عليها من غير أخطاء، فمثلاً عدد نبضات المريض، لنفرض إننا قسنا عدد النبضات فوجدناها 78 فهل يعني هذا بالضرورة أن نبض المريض هو 78؟ فقد يكون الجواب لا لأسباب عديدة منها اختلاف نبض المريض من وقت لآخر تبعاً لحالته الصحية.

4.4 المسح الأولي

يمكن التغلب على كثير من الصعوبات التي تواجهنا في عملية تصميم وتنفيذ المسح بالعينة وحتى المسح الشامل، وذلك من خلال إجراء مسح أولي لبعض عناصر المجتمع قبل أن نقوم بتنفيذ المسح الرئيس للمجتمع. إن المعلومات التي سوف نحصل عليها بهذه الطريقة، يمكن أن تساعدنا لتحديد كثير من العوامل الحرجة التي منها:

1. أخطاء القياسات الممكنة الحدوث.
2. معدل أو نسبة عدم الاستجابة.
3. المواضيع الحساسة أو مصادر عدم الوضوح.
4. مدى الاختلاف بين العادّين.
5. صعوبة الوصول إلى بعض وحدات العينة.
6. مدى التشتت للمتغير الرئيس أو الثانوي.

لا توجد هناك وصفة جاهزة وبسيطة يمكن إعطاؤها لتحديد مقدار المعلومات التي نحتاجها، أو الطريقة التي نستخدمها للحصول على هذه المعلومات الأولية، ويجمع الجميع على ضرورة الحصول على هذه المعلومات والحاجة إليها، ولكن مقدارها يعتمد على مدى وسعة المشكلات الأساسية

التي تحتاج إلى معلومات أولية لها، كذلك تعتمد على مقدار الوقت والمال المتوفرين لها.

نستخدم المعلومات في المسح الأولي في العادة لتعديل وتحسين تصميم المعاينة (على سبيل المثال اختيار حجم العينة)، وتستخدم لتعديل طريقة البحث وطريقة العمل (على سبيل المثال تعديل الأسئلة المطروحة في الاستبانة)، وتعديل التعليمات المعطاة إلى العادّين، والسماح بطرق مختلفة للمتابعة والإشراف، ولا تدخل هذه البيانات التي نحصل عليها هنا ضمن المسح الرئيس فيما عدا بعض طرق المعاينة كالمعاينة المزدوجة أو المعاينة المتكررة.

إن الأداة الرئيسة في عمليات المسح الأولي هي اختبار الموقع Pilot survey ويمكن أن يأخذ عدة أشكال، وذلك يعتمد على الغرض من المعاينة. يمكن أن تقتصر الدراسة على أعداد محدودة جداً من عناصر المجتمع الذين سنجرب معهم الطرق المختلفة للحصول على المعلومات المطلوبة، وهذا النوع مهم جداً ويستخدم كثيراً إذا كان موضوع المسح يتعلق بأمور حساسة أو شخصية أو عندما يكون الموضوع معقداً، ومن المحتمل أن يساء فهم الدراسة. إن إجراء مقابلة واسعة وشاملة وعميقة مع مجموعة من عناصر المجتمع يمكن أن يزودنا بمعلومات أولية مفيدة، يمكن الاعتماد عليها لتعديل تصميم المسح أو الأسئلة المطروحة، أو الطريقة التي تتبع للحصول على المعلومات، حيث إن الأسئلة وطريقة المعاينة وكيفية الوصول إلى الوحدات يمكن أن تدرس وتهذب باستخدام ما يسمى بدراسة الموقع، أو اختبار الموقع لجوانب مختلفة من المجتمع.

يمكن أن نقوم بدراسة الموقع أو المسح الأولي من خلال سحب عينة عشوائية مقبولة الحجم قد يكون حجمها بين 50 إلى 500 وحدة، وغالباً ما يكون هذا المسح الأولي مهماً جداً لنجاح المسح بالعينة الرئيس. ويستخدم هذا

المسح الأولي لتقدير نسبة عدم الاستجابة، كذلك للتخلص من أخطاء الاستجابة المحتملة وبطبيعة الحال سوف يساعدنا في تحديد حجم العينة للمسح الرئيس بشكل أكثر دقة، وأيضًا في تحديد عملية المتابعة، ومن الممكن أن نحصل على الخطوط العريضة للطرق التي يمكن لخطأ الاستجابة أن يحصل بها، وعلاقته بأفراد المجتمع الذي سيؤدي إلى تحيز في التقدير بشكل خاص. فإذا كان عدم الاستجابة أو الإجابة الخاطئة لها علاقة بجزء أو أجزاء من أفراد المجتمع فيمكن أن نستخدم في تقدير التكاليف لأنواع مختلفة من طرق جمع البيانات، وكذلك تقدير تكاليف المسح الرئيس، وأخيراً يمكن استخدام المسح الأولي للحصول على تقديرات أولية لبعض صفات المجتمع، أو متغيرات الدراسة.

إن النقطة الأخيرة هنا في جميع المسوحات سواء كانت أولية أو رئيسة لابد أن نقوم بدراسة ومعرفة كيفية الوصول إلى وحدات المجتمع من قبل العادّين، وهل ستكون الطرق ملائمة، وهل توجد وسائط نقل لنقلهم إلى هذه الوحدات؟ وهل يحتاج العادون إلى إشراف مستمر؟ وهل توجد أجزاء من المجتمع لا يمكن الوصول إليها؟

5.4 طرق جمع البيانات

هنالك عنصران رئيسان في عملية التخطيط للمسح بالعينة، الأول هو اختيار طريقة المعاينة، هل نختار العينة العشوائية البسيطة أو إحدى طرق المعاينة الأخرى التي تكون أكثر تعقيداً، أما الثاني فهو كيف سنقوم بتنفيذ المسح بالعينة بمعنى جمع البيانات وفقاً لطريقة المعاينة التي اخترناها؟ هنالك طرق عديدة لجمع البيانات التي يمكن أن نستخدمها بالاعتماد على طبيعة المجتمع وموضوع البحث أو المسح. سوف نتكلم عن أهم هذه الطرق وهي استخدام المعلومات المتوافرة، وطريقة المشاهدة، والمقابلة وجهاً لوجه،

واستخدام البريد، وأخيراً استخدام الهاتف. وسوف نتطرق إلى كل طريقة من الطرق السابقة على حدة.

1. استخدام المعلومات المتوافرة

أصبح من متطلبات الحياة العصرية اليوم أن نقوم بجمع معلومات عن كل شيء بما فيها المعلومات في كل المجالات، الشخصية، والعلمية والاجتماعية، هذه المعلومات مدونة، ويمكن أن تكون متوافرة على مستويات مختلفة من التفصيل وشروط مختلفة للاستعمال. لقد تم جمع هذه البيانات لأغراض إدارية وحكومية، فمثلاً كشف بأسماء وعناوين كل المشتركين مع دوائر الماء، والكهرباء وأسماء المحلات التجارية والشركات المحلية المختلفة والمسوحات الشاملة التي تقوم بها الدولة كل عشر سنين، وتفاصيل عن كل المركبات المرخصة ... إلخ، وكذلك معلومات أكثر دقة وأكثر شخصية يحتفظ فيها في المستشفيات ودوائر الشرطة والضريبة... الخ، وكذلك تحتفظ هيئات مختلفة بمعلومات عن أعضائها ابتداء بالجمعيات الاجتماعية وانتهاء بالمحامين. وأخيراً يحتفظ الأفراد بمعلومات شخصية عنهم وعن أسرهم.

من الناحية المبدئية كل هذه المعلومات يمكن أن تكون مفيدة ولكن استخداماتها محدودة، وذلك لعدة أسباب:

1. قد لا تكون متطابقة مع المجتمع الذي نهدف إلى دراسته.
2. يمكن أن تكون قديمة.
3. يمكن أن تكون مجمعة مع بعضها بشكل يجعلها غير نافعة.
4. قد لا يمكن الحصول عليها لأسباب قانونية.
5. قد لا تتعاون الجمعيات أو الهيئات مع الباحثين لإعطاء معلومات عن أعضائها بالتفصيل الذي يحتاجونه.

تعد قيمة المعلومات المتوافرة في السجلات محددة على مستوى وحدة المعاينة ما عدا بعض التفصيلات المحدودة، مثل الجنس والعنوان. بعض الأحيان تعد مفيدة إذا أريد إجراء مسوحات باستخدام العينة داخلياً أي ضمن الدوائر والمؤسسات التي تحتفظ بهذه المعلومات، والتي يمكن أن تزودنا بإطار المعاينة، الذي يعد هو الأساس للمسح بالعينة.

2. طريقة المشاهدة

إنه من الطبيعي أن نحصل على معلومات بالملاحظة أو المشاهدة، والذي يجري من دون الحاجة إلى اتصال مباشر بالوحدات. هذا يعني أننا ننظر ولا نسأل. وهذا الأسلوب شائع بين علماء الطبيعة، ولكن قد تكون هذه الطريقة لا تستخدم بشكل منتظم للحصول على بيانات باستخدام المسح بالعينة، ولكن سوف نتكلم عن هذه الطريقة باختصار آخذين في الحسبان المزايا والعيوب لهذه الطريقة.

أما المزايا فتتضمن الموضوعية والدقة وتجنب خطأ الاستجابة. حتى في الاستبانات الاجتماعية نتجنب ردود فعل الناس حول موضوع أو حالة جديدة وقد تكون الملاحظة المباشرة مفيدة، فعلى سبيل المثال في مسح السوق يمكن أن تقوم شركة ما أنتجت بضاعة جديدة بوضعها في منطقة تجريبية، وملاحظة ردود فعل الزبائن حول البضاعة الجديدة من مناطق معينة بدلاً من مقابلة الزبائن وسؤالهم لمعرفة رأيهم في البضاعة الجديدة، والذين قد يتحرجون من الإجابة. إن البيانات التي تجمع بطريقة الملاحظة المباشرة تتحاشى أو تتخلص من مشكلة عدم المعرفة أو عدم فهم السؤال، وفي بعض الأحيان فإن طريقة الملاحظة أو المشاهدة المباشرة لا يوجد بديل عنها، أو هي الأسلوب الوحيد الذي يمكن أن يستخدم، فمثلاً في حالة حشرات تطفو على

سطح ماء النهر لا بد من الذهاب ومشاهدتها، أو حيوانات برية نريد جمع معلومات عنها، أو أسماك في بحيرة، أو ... إلخ.

أما عيوب هذه الطريقة فأهمها: أنها تحتاج إلى وقت كثير وتكاليف عالية، وقد تؤثر في التصرفات الطبيعية لبعض المخلوقات التي نراقبها لجمع معلومات عنها، وقد تكون غير ممكنة التطبيق، فمثلاً لا يمكن أن نتصور أننا نجلس ونراقب أفراداً معينين لمعرفة آرائهم في قضية معينة.

3. المقابلة وجهاً لوجه

تعد المقابلة وجهاً لوجه من الوسائل الشائعة الاستعمال لجمع البيانات، وبصورة خاصة استطلاعات الرأي، حيث يخرج العادون إلى المجتمع مزودين باستبانة مكونة من مجموعة من الأسئلة، ويقومون بمقابلة الأشخاص الذين ظهرت أسماؤهم بالعينة، ويجمعون المعلومات المطلوبة منهم. ولكي يكون المسح ناجحاً يتطلب إعداداً متقناً وتصميماً جيداً لأسئلة الاستبانة والعمل الميداني، وتقيم المسح من خلال المعاينة الأولية، وكذلك إعداد وتدريب العادين والمحافظة على مستواهم، وكذلك التأكيد على مدى قابليتهم لإجراء المقابلات.

إن من أهم مزايا هذا الأسلوب من جمع المعلومات أنه يقلل نسبة سوء الفهم لأسئلة الاستبانة، وكذلك يشجع على زيادة نسبة الاستجابة، وتقليل خطأ عدم الاستجابة. أما الصعوبات التي تواجهنا باستخدام المقابلة فأهمها عدم استجابة الناس للمقابلة، وكذلك يستخدم بعض أفراد العينة أسلوب الإجابة بنعم، أو الموافقة على كل ما يطرح عليهم من أسئلة، وقد يكون هنالك سؤال أو أسئلة مقبولة في المحادثات الخاصة ومرفوضة عندما تكون بصورة رسمية لجمع البيانات، وكذلك هنالك حالات لإعطاء إجابة خاطئة عن قصد أو إعطاء إجابة يعتقد أنها هي المطلوبة. إن الحالة النفسية للعادّين تُعد

من الأمور المعقدة، حيث إنه قد يكون لها تأثيرات كبيرة ومختلفة على جمع البيانات المطلوبة. إن المقابلة الشخصية أو ما نسميها المقابلة وجهاً لوجه مكلفة جداً بالقياس إلى الطرق الأخرى كاستخدام البريد أو الهاتف أو جمع المعلومات من السجلات، ولكن المميزات الكثيرة لهذا الأسلوب والتي من أهمها تقليل نسبة عدم الاستجابة وسوء الفهم أو سوء تفسير السؤال وربما زيادة سرعة الحصول على المعلومات (لأن السرعة قد تكون مطلوبة في حالات استطلاعات الرأي أو أبحاث السوق) يجعل هذه الطريقة شائعة الاستخدام.

لا بد من التأكيد هنا على أن المقابلة الشخصية لا تعني الوقوف على مفترق الطرق لمقابلة الناس، ولكنها في الغالب تجري من خلال تصميم المسح الذي يتضمن اختيار وحدات المعاينة بطريقة عشوائية من المجتمع باستخدام إحدى طرق العشوائية المعروفة، وتحديد أين نجد هؤلاء الأشخاص لمقابلتهم وكيفية الوصول إليهم، وهذا يتضمن تحديد طريقة للوصول إليهم ووسائل النقل التي تستخدم، وكذلك طريقة للتعامل مع الأشخاص الذين لا نجدهم وقت المقابلة، في حالة عدم الاستجابة فإن عملية إعادة المحاولة لمقابلتهم تكون مكلفة.

والمشكلة الأخيرة التي تواجهنا في هذا الأسلوب هي دور العاد نفسه في التأثير على إجابة الأشخاص، وذلك من خلال توضيح الأسئلة أو التعليق على بعض الأمور المطروحة، ومن الناحية النظرية يجب أن يطرح العاد السؤال بموضوعية وبدون أي تحيز، ويمكن أن يعلق على موضوعات جانبية لشرح السؤال إذا لزم الأمر ذلك، وهذا النوع من الخطأ يسمى خطأ العادين، ويمكن أن يكون هنالك تحيز بسبب الاستجابة من قبل الشخص للعاد، وقد يكون هنالك خطر عدم فهم الجواب من قبل العاد، ويمكن التغلب على هذه المشكلة من خلال تدريب ومتابعة العادين.

4. استخدام البريد

إن البديل الواضح والقريب لأسلوب المقابلة هو استخدام البريد، وذلك عن طريق إرسال الاستبانات إلى الأشخاص بالبريد، والطلب منهم الإجابة عنها، وإعادتها إلى الجهة التي أرسلتها لهم، هنالك طريقة مماثلة، وهي أن توزع الاستبانات من قبل العادين على أعضاء العينة وتركها معهم والطلب منهم إعادتها بواسطة البريد، إن توزيع الاستبانات من قبل العادين له ميزات أهمها أن الاتصال الشخصي قد يحث الأشخاص على ملء هذه الاستبانات، وتعطى الفرصة لتوضيح بعض الصعوبات، ويمكن أن ترفع من نسبة الاستجابة، ولكنها مكلفة كما هو الحال بالنسبة إلى المقابلة الشخصية، ولكنها أحيانا تكون أقل كلفة.

عند استخدام طريقة المسح بالبريد يجب أن نراعي أن هناك رغبة لدى الناس لرمي هذا النوع من الرسائل عند استلامها؛ لذلك لا بد من بذل الجهود لمنع مثل هذه الظاهرة كأن تكتب رسالة معدة بشكل جيد ومصممة بعناية مع الاستبانة، وأن يرسل مغلف بريدي عليه طابع وعنوان لإعادة الاستبانة بعد ملئها إلى الجهة التي أرسلتها؛ لأنه من الصعوبة أن ترمي رسالة معنونة عليها طوابع غير مستخدمة.

هنالك مسوحات تقوم بها دوائر الدولة وبعض الشركات الخاصة، وهذا النوع من المسوحات لا يحتاج إلى وسائل لحث عناصر العينة على الرد وخصوصاً إذا كان يجري على أعضاء هذه المؤسسات أو الدوائر؛ لأنه من السهولة أن تطلب هذه الدوائر من العاملين فيها ضرورة ملء هذه الاستبانات وإعادتها إلى الجهة التي طلبتها. ولكن لا يمكن أن نتصور أن هذه الدوائر سوف تحصل على رد كامل وخصوصاً المسوحات التي تقوم بها دوائر الدولة.

تعاني المسوحات في البريد من انخفاض نسبة الاستجابة أو الرد على الرسائل، وقد تصل إلى أقل من 50% لذلك يحتاج الباحثون إلى إرسال رسائل أخرى للتأكيد على ضرورة ملء الاستبانات التي أرسلت إليهم وإعادتها، وقد يتطلب هذا الأمر أكثر من تأكيد، وفي الغالب لا بد أن ترسل مع هذه التأكيدات استبانات جديدة ومغلفات عليها طوابع وعنوان لإعادتها إلى الجهة التي أرسلتها. فوائد أخرى للمسح بالبريد بالإضافة إلى تقليل التكاليف أيضاً التخلص من خطأ العادّين، وإعطاء فرصة لتغطية موضوعات متفرقة بصورة أعمق وأوسع. ولكنها قد تستغرق عدة شهور.

5. استخدام الهاتف

أصبح الهاتف مع تطور الاتصالات موجوداً بشكل كبير في غالبية المنازل وفي متناول الجميع، لذلك ليست فكرة سيئة أن تحاول تنفيذ المسح باستخدام الهاتف والاتصال بالأشخاص أو المؤسسات لجمع البيانات المطلوبة، وفكرة استخدام الهاتف ليست بالجديدة، فلقد استخدمت في الولايات المتحدة عام 1934 لاستطلاع أراء الناس حول بعض الأمور السياسية، ولكنها فشلت لأن حوالي 60% من المنازل في ذلك الوقت لا يوجد فيها تلفون، وهذا واضح لنا لأن المجتمع لم يغط كاملاً، وهنالك نسبة كبيرة من المجتمع لم تكن لها فرصة في الظهور في العينة، وخصوصاً ذوو الدخول المحدودة والذين لا يمتلكون الهاتف. ولكن الأمر اختلف الآن حيث إن أكثر من 90% من المنازل تحتوي على هواتف في الدول المتقدمة وحتى في بعض الدول النامية.

هناك مزايا واضحة لاستخدام الهاتف منها السرعة، وقلة التكاليف، ولكنها لا تخلو من العيوب، وأهمها فيما يتعلق بمسألة تغطية المجتمع، فحتى مع توافر تلفون في 90% من المنازل فإن العينة لا تكون عشوائية، لأن المنازل التي تحتوي على هواتف لها آراء اجتماعية واقتصادية وسياسة قد تكون

مختلفة. لذلك نواجه هنا مشكلة التحيز بسبب عدم تغطية المجتمع كاملاً، ولا بد من تقييمها ومعرفة تأثرها في كل مسح على انفراد.

جميع الصعوبات التي واجهناها في المقابلة الشخصية سوف نواجهها باستخدام الهاتف ولا توجد حاجة إلى إعادتها بالإضافة إلى أن الناس يبدون رفضاً أكبر لاستخدام الهاتف في جمع المعلومات، كما يعتبر البعض هذا تدخلاً في منازلهم أو حياتهم الخاصة سببه الهاتف، لذلك لا بد من استخدام بعض المهارات من قبل العاد الذي يستخدم الهاتف ليتجاوز هذه الصعوبة. على الرغم من المشكلات التي تعزى لهذه الطريقة في جمع المعلومات لكنها تستخدم بصورة واسعة وذلك لسهولتها وقلة تكاليفها وسرعتها.

6. استخدام البريد الإلكتروني

للقيام بالمسح بالعينة باستخدام البريد الإلكتروني، نقوم بتجهيز كشف بعناوين البريد الإلكتروني للأشخاص الذين نرغب في الحصول على معلومات منهم، وهذا ما يسمى بإطار المعاينة، ومن ثم نقوم بسحب عينة عشوائية من بين العناوين الإلكترونية الموجودة في الكشف. ومن ثم نقوم بإرسال رسالة إلى العناوين الإلكترونية التي سحبت بالعينة تحتوي على الاستبانة. هنالك طريقتان لإرسال الاستبانة إما أن ترسل كملحق، أو نضعها ضمن الرسالة الإلكترونية. ونطلب من الشخص أن يجيب عن الأسئلة ومن ثم يقوم بإعادة إرسال الاستبانة إلى نفس العنوان أو إلى عنوان آخر نحدده في نهاية الرسالة. لابد أن هناك رسالة قصيرة تكون ضمن الرسالة الإلكترونية، نوضح للمستجيب في هذه الرسالة من نحن، وما هي أهداف البحث باختصار شديد، مع شرح مبسط لكيفية تعبئة الاستبانة. ولابد أن نؤكد له أن هذه المعلومات سوف تستخدم لأغراض البحث العلمي، أو لمساعدة الإدارات المعنية في اتخاذ القرارات المناسبة، وأن هذه المعلومات لن تستخدم ضده بأي حال من الأحوال.

هناك مجموعة من المصاعب والمشكلات التي تعتري هذا النوع من طرق جمع البيانات نجملها فيما يأتي:

1. هناك صعوبات ومشكلات فنية وتقنية تتعلق بالحاسب الآلي واستخداماته والبرامج المستخدمة مع كل نظام للبريد الإلكتروني وغيرها، فمثلاً في كثير من الأحيان لا نستطيع فتح ملحق معين لسبب أو لآخر، لا أريد الخوض كثيراً في هذا الموضع لأنني لست من المتخصصين في هذا المجال. ولكن لنفترض أن جميع هذه المشكلات تخلصنا منها، وأن المستجيب يستلم الرسالة الإلكترونية ويستطيع الإجابة عنها إلكترونيا.
2. جميع الصعوبات والمشكلات التي واجهتنا في حالة استخدام الهاتف والبريد العادي ستكون منطبقة هنا، فعلى سبيل المثال رفض الشخص للإجابة عن الأسئلة أو تعبئة الاستبانة أو انخفاض عدد المستجيبين (أي نسبة الإجابة قليلة).
3. ينفرد استخدام البريد الإلكتروني ببعض المشكلات الخاصة به منها:

أ. أن كثيراً من الناس إلى الآن لا يستخدمون الحاسب الآلي ناهيك عن استعمال البريد الإلكتروني.

ب. غالباً ما يرفض الأشخاص فتح أي رسالة لا يعرفون مرسلها، ويقومون بحذفها، إما لضيق الوقت أو ظناً منهم أنها رسائل تعلن عن أشياء لا رغبة لهم بها.

ت. من أهم المشكلات التي تواجه استخدام البريد الإلكتروني، هو انتشار الفيروسات عبر البريد الإلكتروني، وما لها من تأثيرات كبيرة على أجهزة الحاسب؛ لذا غالباً ما يقوم الشخص بحذف أي رسالة بريدية لا يعرف مرسلها خصوصاً تلك التي تحمل ملحقاً.

ث. قد لا يفتح الشخص بريده الإلكتروني لأيام عديدة، أو لديه كم هائل من الرسائل بحيث لا يستطيع قراءتها، وغالباً ما يقوم بحذف كثيرٍ من الرسائل خصوصاً تلك التي لا يعرف مرسلها.

ج. من الأخطاء الشائعة في استخدام البريد الإلكتروني أن يقوم الباحثون بإرسال الاستبانة إلى عدد كبير من الأشخاص، أو ربما إلى جميع العناوين الموجودة في الكشف، وهذا سيؤدي إلى أن أعداد غير المستجيبين ستكون كبيرة جداً حتى ولو أرسلنا لهم الاستبانة مرات عديدة.

إن من أهم مزايا استخدام البريد الإلكتروني هي: السرعة، وانخفاض التكاليف وسهولة التنفيذ، وإمكانية الرجوع إلى الوحدة مرات عديدة دون أي تكاليف إضافية، وإمكانية إرسال الرسالة مرة أخرى.

7. استخدام الشبكة العنكبوتية (Internet)

كثيراً ما نرى استبانات أو استطلاعات رأي موجودة في كثير من المواقع على الشبكة العنكبوتية (Internet)، هذه الاستطلاعات لا تمثل سوى رأي المستجيبين ولا يمكن أن نعتبرها مسحاً علمياً يمكن تعميم نتائجه على المجتمع الذي نرغب بمعرفة آراء أفراده أو دراسته.

يمكن الجمع بين طريقة البريد العادي أو الهاتف والشبكة العنكبوتية بأن نقوم بإرسال رسائل إلى الأشخاص الذين سحبوا في العينة أو الاتصال بهم هاتفياً، ونطلب منهم الدخول إلى الموقع في الشبكة العنكبوتية لتعبئة الاستبانة. لابد أن يكون الدخول إلى هذا الموقع أو الجزء الذي يحوي الاستبانة محدوداً بحيث يقتصر على الذين خرجت أسماؤهم في العينة. ويمكننا القيام بذلك بأن نعطي لكل شخص كلمة مرور خاصة به.

جميع العيوب التي تواجه طريقة البريد العادي والهاتف تنسحب على استخدام الشبكة العنكبوتية، يضاف إليها عدم معرفة استخدام الحاسب ومن ثم الدخول إلى الموقع على الشبكة على افتراض أن جميع المشكلات الفنية والتقنية المتعلقة بتعبئة الاستبانة وإرسالها جرى حلها بشكل أو بآخر.

لمزيد من المعلومات حول طرق استخدام البريد الإلكتروني والشبكة العنكبوتية يراجع Malhotra (2006) و Zikmund and Pabin (2006).

6.4 الاستبانة

واضح من خلال مناقشاتنا السابقة أن نجاح أي مسح يتطلب موضوعية في الإجابات من قبل الأشخاص والمؤسسات التي وقع عليهم الاختيار في العينة. وهذا يعتمد بصورة مباشرة على مهارات مصممي الاستبانة وهناك عوامل كثيرة ممكن أن تؤثر في الإجابات عن أسئلة في الاستبانة يتقدمها فهم السؤال، لذلك لا بد أن يكون السؤال واضحاً ودقيقاً، والإجابات المختلفة للسؤال الواحد تلك التي اختيرت أيضا واضحة ومحددة ومختارة بعناية، فمن السهولة قول شيء ولكن من الصعوبة فعله، سوء الفهم قد يأخذ أشكالاً متعددة، فقد يكون السؤال فنياً، أو الموضوع غير مألوف، أو الكلمات غير معروفة. ولنأخذ هذا السؤال:
جهود الحكومة لحفظ التضخم لم تكن ناجحة إلى الآن: يجب أن تعمل شيئاً ما فهل أنت؟

موافق بشدة	موافق	غير موافق	غير موافق بشدة

توجد مشكلات كثيرة في هذا السؤال أولها قد يكون الفرد المسؤول لا يعرف معنى كلمة تضخم، طرح السؤال وجهة نظر سياسة قد لا يريد الفرد أن يقحم نفسه فيها. طرح السؤال بهذه الصيغة قد يفهم منه وقف العمل

بالجهود الحالية أو الاستمرار بالجهود الحالية مع محاولة تعديلها أو تسريعها، وكذلك اختيارات الإجابة عن هذا السؤال لم تعط مجالاً لأن يكون الشخص محايداً أو لا رأي له، وترتيب الإجابات من غير موافق بشدة إلى موافق بشدة قد يكون له تأثير على اختيار الإجابة، وماذا عن مساحة المربعات المخصصة للإجابة حيث إنها غير متساوية المساحة، وقد تؤثر في اختيار الإجابة.

توجد عوامل كثيرة تؤثر في الإجابة منها: أن الأفراد قد يحاولون الإجابة لحماية أنفسهم من أذى قد يصيبهم، أو محاولة الإجابة بشكل يفهم منه أكثر من إجابة. أو أنهم يجيبون لتوقعهم أن السائل يريد منهم إجابة معينة، لذلك لا بد من صياغة الأسئلة والإجابات بشكل يقلل كثيراً من هذه الصعوبات.

جانب آخر هو الأسئلة الشخصية أو الحساسة التي قد لا يرغب الفرد في الإجابة عنها، لأسباب كثيرة فمثلاً إذا سئل عن دخله قد يخشى أن نفرض عليه ضريبة أو قد يخشى الحسد، كثيرٌ من الفلاحين إذا سألناهم عن مقدار الحليب الذي يحصل عليه من بقراته، قد لا يعطينا الإجابة الصحيحة، وللتغلب على هذه المشكلات فيما يخص هذا النوع من الأسئلة يعمد إلى أسلوب الأسئلة غير المباشرة فبدلاً من أن نسأل الشخص عن دخله الشهري فمن الممكن أن نسأله عن نوع المنزل الذي يسكنه، وعن المنطقة التي هو ساكن فيها وعن قراءته وعاداته. إن هذه الأمور في مجملها يمكن أن تعطينا فكرة عن دخله، ويمكن أن نستعمل أسلوب الإجابة العشوائية على السؤال فمثلاً لتقدير معدل عدد السرقات بالنسبة لعدد أفراد المجتمع يمكن أن نسأل: هل ولدت في شهر معين أو سجنت في السنة الماضية بسبب السرقة، بالإضافة إلى هذه الأمور التي تكلمنا عنها، وهناك عوامل أخرى يمكن أن تؤثر في الإجابات ومنها: موضع السؤال في الاستبانة، وعلاقته بالأسئلة

الأخرى، والكلمات التي صيغ بها السؤال، واهتمام الأشخاص بالموضوع المطروح، والحاجة إلى الذاكرة في الإجابة عن السؤال.

إن تصميم الاستبانة موضوع مهم جداً وله تأثير كبير في نتائج المسح، لذا يجب أن يعطى أهمية كبيرة. لمزيد من المعلومات راجع: Grover (1989) و Schuman and Presser (1981) و Homville and Jowell (1978) و Gillham (2000) و Moser and Kalton (1986) و Robson (1993).

7.4 الخطوات الأساسية لتجهيز البيانات

تجميع البيانات وفقاً لطريقة المعاينة المناسبة التي اختيرت لهذا الغرض يُعدُّ أمراً جيداً. ولكن السؤال الآن ماذا نعمل بهذه البيانات؟ الجواب الأولي نستعملها لتقدير معلمات المجتمع مع تقدير دقتها وإجراء بعض الاختبارات الإحصائية، ولكن قبل الانتقال إلى هذه المرحلة هناك بعض القرارات الأساسية عن تنظيم وتسجيل البيانات لا بد من اتخاذها ليتسنى تحليلها، مع التأكيد على الحاجة إلى الحاسب الآلي والحقائب الإحصائية الجاهزة، التي يمكن استخدامها لأغراض تنظيم وتسجيل وتحليل البيانات، فاستخدام الحاسب في أيامنا هذه أصبح ضرورة قصوى في غالبية أعمالنا، ومنها استخدام الحاسب لحفظ البيانات وتحليلها، وفي الحقيقة فإن الكثير من المسوحات لا يمكن إنجازها بدون الحاسب، وبالنسبة إلى تجهيز البيانات فيمكن أن تمر بمراحل هي: تدقيق البيانات، وتجهيزها لكي تدخل في الحاسب، ومن ثم تحليلها وعرضها، ولاستخدام الحاسب أصبح ضرورياً أن ترقم الإجابات، وذلك حتى لا نحتاج إلى طاقة تخزينية كبيرة. فمثلاً السؤال عن الجنس وجوابه ذكر أو أنثى يمكن أن نستخدم 0 و 1، هذه الأرقام التي استخدمت لا معنى لها سوى أن الصفر يرمز إلى الذكر والواحد يرمز إلى

الأنثى، ويمكن أن نستخدم أي رقمين آخرين. كذلك السؤال حول المستوى التعليمي الذي وصل إليه الشخص بحيث يمكن أن يكون جوابه.

أمي
يقرأ ويكتب
حاصل على التوجيهي
حاصل على درجة جامعية
دراسات عليا

ويمكن أن ترمز إلى هذه الإجابات المختلفة 4,3,2,1 ,0 يبدو أن الترقيم هنا له معنى وهو ترتيب طبيعي لمستوى التعليم حيث الصفر يرمز إلى " أمي" و 1 يرمز إلى "يقرأ ويكتب" وهكذا.

نقوم بترقيم أو ترميز جميع إجابات الأسئلة الواردة في الاستبانة. وتكون عملية الترقيم هذه هي البداية في تحليل البيانات وهي في الغالب كذلك. لكن الخطوة التي تأتي مباشرة بعد الحصول على البيانات هي عملية تدقيق البيانات، والتأكد من صحة الإجابات، ومحاولة التخلص أو تصحيح بعض الأرقام أو تعديلها، فمثلاً من غير المعقول أن يكون عمر شخص 500 سنة، على الأرجح إن عمره 50 سنة، وبعد التأكد من سلامة البيانات من الأخطاء تبدأ عملية تجهيز البيانات لإدخالها إلى الحاسب الآلي وعملية التجهيز هذه تعد ضرورة لضمان سير عمليات الإدخال بصورة سريعة ودقيقة وبدون أخطاء، ولكن أعمال التدقيق لا تقتصر على عملية التدقيق قبل إدخال البيانات في الحاسب، ولكن تستمر بعد الإدخال للتأكد من سلامة البيانات من أخطاء الإدخال أو حتى بعض الأخطاء التي لم تستطع اكتشافها قبل إدخال البيانات في الحاسب. ويمكن القيام ببعض هذه الأعمال باستخدام الحاسب.

لقد وجهت عناية خاصة في السنوات الأخيرة للتعامل مع الإجابات الخاطئة وذلك إحصائياً، ومن خلال محاولة توقُّع الإجابة الممكنة، ومن ثم محاولة مقارنة الأرقام أو المعلومات الموجودة باستخدام الحاسب إحصائياً أو روتينياً للتأكد من صحة الإجابات. وأخيراً لا بد من الإشارة هنا إلى أن أي تحليل تفصيلي للبيانات يحتاج إلى وضع البيانات في جداول بسيطة يسهل فهمها والتعامل معها، وتأخذ أشكالاً مختلفة من أهمها جداول التوزيعات التكرارية للإجابة عن كل سؤال، وكذلك جدول التوزيع الثنائي للإجابة عن سؤالين.

تمارين

1. يرغب أحد الباحثين في تقدير معدل المصروفات الشهرية من المياه لإحدى المدن. ناقش فكرة اختيار وحدة المعاينة، وكذلك ناقش كيف يمكن الحصول على إطار للمعاينة؟
2. تنوي دائرة الغابات في إحدى المناطق تقدير نسبة الأشجار التي يزيد قطرها عن 15 سم. إذا كان متوفراً لدى الدائرة خارطة بجميع الغابات الموجودة في المنطقة، ناقش فكرة وحدة المعاينة وكيف يمكن الحصول على إطار للمعاينة؟
3. يحتوى أحد المجتمعات على 8940 وحدة قمنا بترقيمها من 0001 إلى8940. قمنا بسحب عينة عشوائية بحجم 15 وحدة ومن ثم اشرح الطريقة التي استخدمتها لسحب العينة.
4. يرغب مراقب السير في إحدى المدن تقدير نسبة الإطارات التي تُعد غير صالحة للاستعمال والمستخدمة من قبل سائقي السيارات، هل تعتقد أنَّ وحدة المعاينة يجب أن تكون إطارًا أو سيارة أو مجموعة من السيارات واقفة في موقف خاص للسيارات. ما هو إطار المعاينة؟
5. ناقش الطريقة أو الطرق التي يمكن استخدامها لجمع البيانات في الحالات الآتية:

أ. ترغب إحدى مؤسسات الإنتاج التلفزيوني في تقدير نسبة المشاهدين الذين يراقبون البرامج المنتجة
من قبلهم في إحدى البلدان.

ب. ترغب إدارة إحدى الصحف المحلية معرفة رأي قرائها حول تغطيتها الأخبار المحلية والعالمية.

ت. تود دائرة الصحة العامة تقدير نسبة الكلاب التي ستقوم بمهاجمة المارة خلال الشهر القادم في
أحد الأحياء السكنية.

6. يعتقد كثير من الناس أن السياسة المستخدمة للقبول في كثير من الجامعات العالمية، والتي تعتمد على علامات الطلاب في الدراسة الثانوية غير دقيقة، وقد تكون غير عادلة في حق كثير من الطلبة، صمم استمارة لدراسة سياسة القبول التي يمكن أن تكون بديلة عما هي مستخدمة الآن.

7. سئل خريجو طلبة الثانوية العامة في إحدى الاستبانات: ما هو المساق الذي درست في الثانوية العامة، والذي كان له الدور الكبير في تحديد مستقبلك الدراسي؟ فكانت إجابة 25% من الطلبة رياضيات و25% اللغة الإنجليزية. هل تعتقد أنَّ هذا السؤال جيد، وأنَّ نتائجه يمكن الاعتماد عليها؟

8. ناقش ما هي الصعوبات التي سنواجهها في الحصول على إطار للمعاينة للحالات الآتية:

أ. إذا أردنا دراسة مصروفات طلبة جامعة الملك فهد للبترول والمعادن الشهرية.

ب. إذا أردنا دراسة الحيوانات البرية الموجودة في صحراء الربع الخالي.

ت. إذا أردنا دراسة مستوى معيشة الفلاحين في غور الأردن.

ث. إذا أردنا دراسة نسبة الفواتير المتأخرة الدفع إلى شركة الكهرباء في مدينة الدمام.

9. لمعرفة رد فعل طلبة إحدى الجامعات حول طرح مساق جديد بالحاسب كمتطلب جامعي لجميع طلبة الجامعة، صمم استمارة استبانه لكي نقوم

بتوزيعها على طلبة الجامعة لكي نستخدم المعلومات التي حصلنا عليها من هذه الاستبانة لتصويب تدريس هذا المساق في الجامعة.

10. صمم استمارة استبانه لجمع معلومات لدراسة الحالة الاجتماعية لطلبة في إحدى كليات الجامعة، تتضمن محل الميلاد، والمشكلات التي يواجهها الطالب في الحصول على مسكن في المدينة، وكيفية قضاء أوقات فراغه، والهوايات التي يمارسها، ومصروفاته الشهرية (موزعة على السكن والغذاء والكتب ... إلخ).

11. ما هو رأيك في صحة تمثيل العينات الآتية للمجتمعات التي سحبت منها:

أ. أراد أحد الباحثين دراسة الحالة الاجتماعية لطلبة جامعة اليرموك فقام بالوقوف عند البوابة الرئيسة للجامعة الساعة الثامنة صباحاً، وسأل بعض الطلاب الذين دخلوا الجامعة بعض الأسئلة حول أحوالهم الاجتماعية.

ب. أراد أحد الباحثين دراسة الحالة الاقتصادية لعمال أحد المعامل، حصل على كشف بأسماء العمال، وقام باختيار العمال الذين يحملون الأرقام 6, 26, 46.

ت. أراد أحد الأطباء العاملين في مجمع للصناعات الكيماوية دراسة الحالة الصحية لعمال المجمع، فجاء بجهاز للأشعة السينية وفحص من تقدم من العمال للفحص، وأعدَّ تقريراً عن الحالة الصحية لعمال المجمع.

ث. لدراسة أحوال الفلاحين في إحدى القرى، ذهب الباحث إلى مختار القرية وسأله مجموعة أسئلة عن بعض العائلات الساكنة في هذه القرية.

12. يطلب من الطلبة اختيار عينة عشوائية واعتباطية من طلبة الفصل لدراسة معدل أطوال الطلبة في الصف، ثم يقارن بين النتائج التي حصل عليها كل طالب، والنتائج الحقيقية لطلبة الفصل.

المراجع العربية

1. وليم كوكوران – ترجمة أنيس كنجو – تقنية المعاينة الإحصائية (1995)، جامعة الملك سعود ، الرياض.
2. حسين علوان (1994) ، طرق المعاينة ، دار الفرقان، عمان.

References

1. Barnett, V. (1991). Sampling Survey Principles and Methods, Edward Arnold, London.
2. Cochran, W. G. (1977). Sampling Technique, 3rd Ed. John Wiley & Sons, New York.
3. Fisher, R. A. (1953). The Design of Experiments, Oliver & Boyd, London.
4. Govindarajulu, Z (1999). Elements of Sampling Survey and Methods, Prentice Hall, New Jersey.
5. Hajek, J (1981). Sampling from Finite Population, Marcel Dekker, New York.
6. Lahiri, D. B. (1951). A Method of Sample Selection Providing Unbiased Ratio Estimates Bull. Int. Statist., 33, 133, 140.
7. Gillham, Bill (2000). Development of Questionnaire, Continuum, London.
8. Groves, R. M (1998). Nonresponse in a Household Interview Survey, Wily, New York.
9. Lavrakas, P.J. (1993). Telephone Survey Methods□Sampling Selection and Supervision, 2^{nd}, Sage, Newbury Park.
10. Lessler, J. T., and Kalsbeek, W. D. (1992). Nonsampling Error in Surveys, Wily, New York.
11. Levy, P. S. and Lemeshow S (1991). Sampling of Population□Methods and Applications, Wiley, New York.
12. Mahalanobis P. C. (1944). On the Large□Scale Sample Surveys, Phil. Trans. R. Soc., 231, 329□451.
13. Mahalanobis, P. C. (1950). Cost and Accuracy of Results in Sampling and Complete Enumeration, Bull. Int. Statist. Inst., 32, 210□213.
14. Malhotra, N. K. (2006). Marketing Research An Applied Orientation, 5 Ed. Prentice Hall, New Jersey.
15. Moser, C. A. and Kalton, G. (1986). Survey Methods in Social Investigation, Aldershot□ Gower Publisher, London.
16. Neyman, J. (1959). Bias in Survey due to Nonresponse. In New Developments in Survey Sampling (Johnson, N. L. and Smith H. Eds.) p 712□732, Wiley, New York.
17. Oppenheim, A. N. (1992). Questionnaire Design, Interviewing and Attitude Measurements, New ed, Printer Publisher, London.
18. Robson, C. (1993). Real World Research□A Resource for Social Scientists and Practitioner Researchers, Blackwell, Oxford.
19. Sampath, S. (2001). Sampling Theory and Methods, Alpha Science International Ltd. U. K.
20. Scheaffer, R. L., W. Mendenhall and L. Ott (1996). Elementary Survey Sampling, 5^{th} Ed., Duxbury press, Boston.
21. Schuman, H., and Presser, S. (1981). Questions and Answers in Attitude Surveys, Academic Press New York

22. Seal, K. C. (1962). Use of Outdated Frames in Large Sample Surveys, Bull□Cal. Statist Assoc., 11, 68□84.
23. Singh, D. (1978). Taking Agricuitural Censuses, FAO, Rome.
24. Sturat A. (1984). The Ideas of Sampling, Revised Edition, Griffin, London.
25. Thompson, S. K (2002). Sampling, 2end Wiley, New York.
26. Yates, F. and P. M. Grundy, (1953). Selection without Replacement from Within Strata Probability Proportional to Size J. R. Statist, Soc. 15 B, 253□261.
27. Yates, F. (1960). Sampling Methods for Censuses and Surveys, Charles Grittin & Co., London.
28. Yates, F. (1981). Sampling Methods for Censuses and Surveys 4th Ed., Grittin, London.
29. Zarkovich, S. S. (1961). Sampling Methods and Censuses, FAO, Rome.
30. Zikmund, W. G. and Pabin, B. J. (2006). Exploring Marketing Research, 9 Ed., South□ Western College Pub.

الفصل الخامس

العينة العشوائية البسيطة
Simple Random Sample

1.5 مقدمة

هناك أساليب كثيرة لاختيار عينة عشوائية بسيطة بحجم n من الوحدات من مجتمع يحتوي على n من الوحدات، ومهما كانت التحضيرات الأولية للأعداد والتحضير إلى المسح بطريقة العينة العشوائية البسيطة لابد من سحب هذه العينة من المجتمع بطريقة أو بأخرى.

تعرف العينة العشوائية البسيطة على أنها الطريقة التي نسحب بها n وحدة من وحدات المجتمع التي عددها N وحدة بحيث تكون أي عينة بالحجم n لها الفرصة نفسها بأن تكون هي العينة المسحوبة. إذا كان المجتمع يحتوي على Nمن الوحدات وأردنا أن نسحب عينة حجمها n بدون إرجاع سيكون لدينا

$$\binom{N}{n} = \frac{N!}{n!(N-n)!}$$

عينة ممكنة للسحب، وكل عينة متميزة عن الأخرى، ونعني بالمتميزة أن لا تكون جميع عناصر أي عينة مشابهة لجميع عناصر عينة أخرى، على سبيل المثال، إذا كان لدينا مجتمع يحتوي على 5 عناصر أي N=5 ونريد أن نسحب عينة حجمها 2 أي إن n=2 فيكون لدينا 10 عينات ممكنة السحب أي:

$$\binom{5}{2} = \frac{5!}{2!(5-2)!} = 10$$

كيف يمكن سحب عينة واحدة بحجم وحدتين من بين العشر عينات الممكنة الظهور؟ إنه من حسن الحظ أن نستطيع سحب هذه العينة، وذلك عن طريق استخدام جداول الأعداد العشوائية أو الحاسب الآلي، وذلك بسحب وحدة واحدة كل مرة من المجتمع وبدون إرجاع إلى أن نسحب n=2 من الوحدات، وهذا يمثل العينة المطلوبة. ولكن السؤال هو هل احتمال سحب هذه العينة مساو لاحتمال سحب جميع العينات الممكنة السحب؟ أثبت Raj (1968) أن العينة التي تسحب وحداتها كل مرة وحدة من المجتمع وبدون إرجاع يكون احتمال سحبها مساوياً لجميع العينات الأخرى الممكنة السحب.

مثال 2.5

لنفترض أن لدينا عائلة مكونة من 4 أفراد، وهذه العائلة تمثل المجتمع الذي ننوي دراسته، وبالتحديد ننوي دراسة أطوال هذه العائلة، فإذا كانت هذه العائلة مؤلفة على وجه التحديد من الأشخاص: أ، ب، ج، د، وقمنا بقياس أطوالهم فوجدناها حسب الترتيب110، 130، 150، 170، تكون البيانات كما يلي:

(أ ، 110) (ب، 130) (ج، 150) (د ، 170)

إن معدل أطوال المجتمع (العائلة) 140 لنفرض أننا نريد أن نسحب عينة حجمها n=2 سيكون لدينا عدد العينات الممكنة السحب6، الجدول الآتي يمثل كل العينات الممكنة السحب.

جدول 1: العينات الممكنة السحب بحجم 2 سحبت من مجتمع عدد وحداته4

رقم العينة	1	2	3	4	5	6
عناصر العينة	(ا، ب)	(ا، ج)	(ا، د)	(ب، ج)	(ب، د)	(ج، د)
قيم العينة	(110 ،130)	(110 ،150)	(110 ،170)	(130 ،150)	(130 ،170)	(170 ،150)

لو أننا قمنا بترقيم هذه العينات من 1 إلى 6 كما هو واضح في الجدول، وقمنا بسحب رقم واحد من بين الأرقام الستة لكان احتمال ظهور كل عينة هو $\frac{1}{6}$ أي إن احتمال سحب كل عينة ممكنة مساو إلى احتمال سحب العينات الأخرى، ولكن كما هو واضح لا يمكن تطبيق هذه الطريقة لصعوبتها من الناحية العملية، على سبيل المثال: إذا كان حجم المجتمع $N=100$ وحجم العينة التي ننوي سحبها $n=2$ فإن عدد العينات الممكنة الظهور أو السحب هي 4950 عينة.

نلاحظ أنه من غير الممكن أن نكتب جميع العينات الممكنة الظهور. لذلك تم الاستعاضة عن هذه الطريقة بأن نقوم بسحب وحدة واحدة كل مرة من المجتمع وبدون إرجاع. قد يسأل سائل ما هو الغرض إذن من هذا المثال غير الواقعي؟ حيث إنه لا يوجد مجتمع بهذا الصغر ونريد دراسته عن طريق استخدام العينات، للجواب عن هذا التساؤل شقان: الأول هو توضيح الفكرة الأساسية وراء سحب العينة العشوائية البسيطة، والسبب الثاني وهو الأهم أننا سوف نستخدم هذا المثال في هذا الفصل لتوضيح بعض الأفكار الأساسية والمهمة والتي تستخدم كثيراً في طرق المعاينة.

3.5 رموز ومصطلحات

N: عدد وحدات المجتمع الذي ننوي دراسته

n: عدد وحدات العينة التي سحبت من المجتمع

y_i: قيمة المتغير y للوحدة i من وحدات المجتمع

$y_1, y_2, ..., y_N$: قيم المتغير y لجميع وحدات المجتمع

$y_1, y_2, ..., y_n$: قيم المتغير y لوحدات العينة

$Y=\sum_{i=1}^{N} Yi$: مجموع المتغير y في المجتمع

$\bar{Y}=\frac{1}{N}\sum_{i=1}^{N} Yi$: الوسط الحسابي للمجتمع للمتغير y

$S^2=\frac{1}{N-1}\sum^{N}(y_i-\bar{Y})^2$: تباين المجتمع للمتغير y

A: عدد وحدات المجتمع التي تحمل صفة مميزة أو تنتمي إلى فئة معينة

$P=\frac{A}{N}$: نسبة الوحدات في المجتمع التي تحمل صفة مميزة أو تنتمي إلى فئة معينة

4.5 تقدير الوسط الحسابي للمجتمع

كما هو متوقع فإن تقدير الوسط الحسابي للمجتمع $\bar{Y}$ بالاعتماد على البيانات التي سحبت باستخدام العينة العشوائية البسيطة هو الوسط الحسابي للعينة:

$$\bar{y}=\frac{1}{n}\sum_{i=1}^{n} y_i$$

إن من أهم صفات الوسط الحسابي للعينة $\bar{Y}$ أنه تقدير غير متحيز إلى الوسط الحسابي للمجتمع $\bar{Y}$ أي:

$$E(\bar{y})=\bar{Y}$$

حيث إن E ترمز إلى التوقع الرياضي.

إن البرهان الرياضي لهذه الميزة ليس بصعب، ولكننا سوف لا نستخدم الأسلوب الرياضي لبرهان هذه الميزة المهمة والميزات الأخرى، يراجع Cochran(1977) أو ترجمة أنيس كنجو(1995) لهذا البرهان الرياضي،

وكذلك جميع البراهين الرياضية الأخرى. وبدلاً من الأسلوب الرياضي سوف أستخدم المثال 2.5 لشرح فكرة عدم التحيز للوسط الحسابي للعينة $\bar{y}$. وذلك من خلال دراسة توزيع لكل العينات الممكنة الظهور. وفي الحقيقة عدم التحيز بالنسبة للوسط الحسابي للعينة$\bar{y}$ يعني إذا أخذت جميع العينات الممكنة الظهور أو السحب بحجم n من المجتمع الذي يحتوي على N من الوحدات فإن الوسط الحسابي لأوساط جميع العينات الممكنة الظهور يجب أن يساوي الوسط الحسابي للمجتمع $\bar{Y}$ أي

$$Aver(\bar{y}) = \frac{1}{\binom{N}{n}} = \sum_{i=1}^{\binom{N}{n}} \bar{y}_i = \bar{Y}$$

رقم العينة	1	2	3	4	5	6
$\bar{y}_i$	120	130	140	140	150	160

$$Aver(\bar{y}) = \frac{1}{6} = \sum_{i=1}^{6} \bar{y}_i = \frac{840}{6} = 140$$

أما الوسط الحسابي للمجتمع:

$$\bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i = \frac{1}{4}\sum_{i=1}^{4} y_i = \frac{840}{6} = 140$$

واضح أن معدل جميع الأوسط الحسابية للعينات الممكنة الظهور مساو للوسط الحسابي للمجتمع $\bar{Y}$. أي أن $\bar{y}$ تقدير غير متحيز إلى $\bar{Y}$. الآن ننتقل إلى تباين $\bar{y}$ حيث إنه يعرف:

$$V(\bar{y}) = \frac{N-n}{n}\frac{S^2}{n} = \frac{1-f}{n} S^2$$

حيث إن $f = n/N$: يسمى كسر المعاينة، وكذلك $(1-f)$ يسمى معامل التصحيح للمجتمعات المحدودة أو الصغيرة الحجم. بإمكاننا إهمال معامل التصحيح للمجتمعات المحدودة إذا كانت قيمته أقل من 5%.

5.5 تقدير تباين المجتمع

يمكن تلخيص استعمالات تباين الوسط الحسابي للعينة $V(\bar{y})$ بما يأتي:

1- للحصول على دقة تقدير $\bar{y}$ للوسط الحسابي للمجتمع $\bar{Y}$.

2- لمقارنة $\bar{y}$ بتقديرات أخرى للوسط الحسابي للمجتمع $\bar{Y}$.

3- لتحديد حجم العينة التي نحتاجها للوصول إلى الدقة المطلوبة لتقدير الوسط الحسابي للمجتمع $\bar{Y}$ باستخدام الوسط الحسابي للعينة $\bar{y}$.

ليس من السهل معرفة S^2 لذلك لابد من تقديرها من العينة العشوائية البسيطة التي سحبت من المجتمع وذلك باستخدام ما يسمى بتباين العينة s^2 والذي يعرف:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

إن تباين العينة s^2 تقدير غير متحيز إلى تباين المجتمع S^2 ، والآن نعود إلى مثالنا في 2.5 لإثبات هذه الخاصية. إذا وجدنا تباين كل عينة ممكنة السحب والتي عددها 6 عينات. وأوجدنا الوسط الحسابي لهذه التباينات فإنه يساوي تباين المجتمع أي:

رقم العينة	1	2	3	4	5	6
s_i^2	200	800	1800	200	800	200

الآن

$$Aver(s^2) = \frac{1}{\binom{N}{n}} \sum_{i=1}^{\binom{N}{n}} s_i^2 = \frac{200+800+...+200}{6} = 666.67$$

أما تباين المجتمع

$$S^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - \bar{Y})^2 = 666.67$$

وهذا يعني أن s^2 تقدير غير متحيز إلى S^2.

6.5 تقدير تباين الوسط الحسابي للعينة

لقد سبق أن عرفنا تباين $\bar{y}$ بأنه:

$$V(\bar{y}) = \frac{1-f}{n} S^2$$

ولكن تباين المجتمع S^2 غير معروف إذاً لابد من استعمال s^2 تباين العينة الذي هو تقدير غير متحيز إلى S^2 ليصبح تقدير تباين الوسط الحسابي للعينة $\bar{y}$:

$$s^2(\bar{y}) = \frac{1-f}{n} s^2$$

واضح أن $s^2(\bar{y})$ تقدير غير متحيز إلى $V(\bar{y})$ ، لإثبات هذه الخاصية لنعد إلى مثالنا في 2.5 لنجد الجدول الآتي:

رقم العينة	1	2	3	4	5	6
$\bar{y}_i$	120	130	140	140	150	160
$(\bar{y}_i - \bar{Y})^2$	400	100	0	0	100	400

$$V(\bar{y}) = \frac{1}{\binom{N}{n}} \sum_{i=1}^{\binom{N}{n}} (\bar{y}_i - \bar{Y})^2 = \frac{1}{6}(1\,000) = 166.67$$

أما تباين $\bar{y}$:

$$V(\bar{y})=\frac{1-f}{n}S^2=\frac{1-2/4}{2}(666.67)=166.67$$

وهذا يثبت أن $s^2(\bar{y})$ تقدير غير متحيز إلى $V(\bar{y})$. ولكن في غالبية كتب الإحصاء يعرف تباين الوسط الحسابي للعينة بما يأتي:

$$V(\bar{y})=\frac{s^2}{n}$$

حيث إن:

$$s^2=\frac{1}{N}\sum_{i=1}^{N}(y_i-\bar{Y})^2$$

لو استخدمنا البيانات الموجودة في مثال 2.5 لوجدنا

$$s^2=\frac{1}{N}\sum_{i=1}^{N}(y_i-\bar{Y})^2=500$$

لذلك فإن تباين $\bar{y}$

$$V(\bar{y})=\frac{s^2}{n}=\frac{500}{2}=250$$

والسؤال الآن لماذا هذا الاختلاف بين قيمة $V(\bar{y})$ في الحالتين؟ في الحقيقة إن المعادلة $V(\bar{y})=\frac{s^2}{n}$ وجدت للمجتمعات غير المحدودة أو الكبيرة جداً، ولكن في مثالنا هنا المجتمع محدود وصغير جداً لذلك نجد هذا الاختلاف الواضح. وسوف نستخدم في هذا الكتاب

$$V(\bar{y})=\frac{1-f}{n}S^2$$

حيث إن

$$S^2 = \frac{N}{N-1}s^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{Y})^2$$

7.5 فترة ثقة للوسط الحسابي

لإيجاد فترة ثقة للوسط الحسابي $\bar{Y}$ يستخدم الوسط الحسابي للعينة $\bar{y}$ لهذا الغرض، ولإيجاد هذه المدة لابد من معرفة توزيع $\bar{y}$ ، هناك حالتان لمعرفة توزيع $\bar{y}$:

1. إذا كانت $y_1, y_2,..., y_n$ ، تتبع التوزيع الطبيعي. فإن $\bar{y}$ كذلك يتبع التوزيع الطبيعي أي:

$$\bar{y} \sim N(\bar{Y}, \frac{1-f}{n}S^2)$$

2. إذا كان حجم العينة $n < 30$ فإننا نستطيع أن نستخدم نظرية تقارب التوزيعات الاحتمالية لنجد أن $\bar{y}$ يتبع التوزيع الطبيعي التقريبي أي:

$$\bar{y} \sim N(\bar{Y}, \frac{1-f}{n}S^2)$$

إذا كانت S^2 معلومة فإن فترة $(1-a)100\%$ ثقة الوسط الحسابي $\bar{Y}$ يمكن أن تعرف:

$$\bar{y} - z_{a/2}S\sqrt{\frac{1-f}{n}} \leq \bar{Y} \leq \bar{y} + z_{a/2}S\sqrt{\frac{1-f}{n}}$$

حيث إن $Z_{a/2}$ يمكن إيجادها باستخدام جدول التوزيع الطبيعي المعياري بحيث

$$P(Z > z_{a/2}) = a/2$$

إذا كان $\bar{y}$ تقريباً يتبع التوزيع الطبيعي ففترة الثقة للوسط الحسابي أعلاه تكون تقريبية. وإذا كانت S^2 غير معلومة نستخدم بدلاً منها s^2 ومن ثم نستخدم توزيع t بدلاً من توزيع z لتصبح فترة الثقة للوسط الحسابي كما يأتي:

$$\bar{y}-t_{n-1,a/2}s\sqrt{\frac{1-f}{n}} \le \bar{Y} \le +t_{n-1,a/2}s\sqrt{\frac{1-f}{n}}$$

حيث إن (n) تسمى درجات الحرية، ونستخدم جدول توزيع t لإيجاد قيمة $t_{n-1,a/2}$.

مثال (1): لمعرفة معدل عدد أيام الإجازة الإضافية التي يأخذها عمال مجموعة من المصانع عند نهاية إجازة عيد الفطر وعيد الأضحى، سحبت عينة عشوائية حجمها 100 عامل من 36000 عامل يعملون في هذه المصانع، وسُئلوا عن عدد أيام الإجازة الإضافية التي أخذوها بسبب إجازة عيد الفطر أو الأضحى فكانت إجاباتهم كما يلي:

y_i عدد الأيام	0	1	2	3	4	5	6	7	8
f_i عدد العمال	45	17	16	10	6	2	1	3	0

لتقدير معدل عدد الأيام التي يأخذها العمال إجازة إضافية بسبب إجازة العيدين نستخدم الوسط الحسابي للعينة $\bar{y}$ لنجد أن:

$$\bar{y}=\frac{1}{n}\sum_{i=1}^{k} y_i f_i=\frac{1}{100}(1\,400)=1.4$$

مع تباين عينة $s^2=3.071$. بما أن حجم العينة كبير فإن $\bar{y}$ تقريباً يتبع التوزيع الطبيعي لذلك فإن فترة 95% ثقة للوسط الحسابي $\bar{Y}$ تكون:

$$1.05701 \langle \bar{Y} \langle 1.74299$$

إذا أهملنا معامل التصحيح تصبح الفترة:

$$1.05653 \langle \bar{Y} \langle 1.74347$$

8.5 تقدير المجموع الكلي للمجتمع

هنالك حالات كثيرة يكون فيها تقدير مجموع المجتمع الكلي لمتغير معين أكثر أهمية من المعدل، على سبيل المثال مجموع أوزان الحقائب التي سيتم شحنها على متن الطائرة أهم بكثير من معدل وزن الحقيبة الواحدة.

ويمكن تعريف مجموع المجتمع الكلي:

$$Y = \sum_{i=1}^{N} y_i = N\overline{Y}$$

باستخدام نتائج العينة العشوائية البسيطة، وحيث إن N حجم المجتمع معروف فإن:

$$Y = N\bar{y}$$

تقدير غير متحيز إلى المجموع الكلي للمجتمع Y وتباين Y :

$$V(Y) = N^2\, V(\bar{y}) = N^2 \frac{1-f}{n} S^2$$

بما أن S^2 غير معروفة فإننا نستخدم s^2 تباين العينة ليصبح تقدير تباين Y :

$$s^2(Y) = N^2\, s^2(\bar{y}) = N^2\, \frac{1-f}{n} s^2$$

كذلك تقدير الخطأ المعياري يصبح:

$$s(Y) = \sqrt{s^2(Y)}$$

إذا كان حجم العينة أكبر من 30 وحدة فإن Y يتبع التوزيع الطبيعي التقريبي أي:

$$\cdot\ N\left(Y, N^2 \frac{1-f}{n} S^2\right)\ Y$$

وبالطريقة نفسها يمكن إيجاد فترة $(1-\alpha)100\%$ ثقة تقريبية للمجموع الكلي Y:

$$Y \mp z_{\alpha/2} N S \sqrt{(1-f)/n}$$

إذا كانت S^2 غير معروفة نستخدم s^2 بدلاً منها و $t_{n-1,\alpha/2}$ بدلاً من $z_{\alpha/2}$.

9.5 تقدير النسبة

لنفرض أنه يوجد لدينا حالة يهمنا فيها فقط إذا كان المتغير y ينتمي إلى فئة معينة أو يتصف بصفة معينة، فمثلاً في صف يتكون من 40 طالباً يهم المدرس أن يعرف ما نسبة أو عدد الطلاب الراسبين في الامتحان الأول، أو في مصنع يهم قسم مراقبة جودة الإنتاج إذا كانت الوحدة المصنوعة مطابقة للمواصفات أم لا. لذلك نستطيع أن نعيد تعريف المتغير y ليصبح:

إذا كان المتغير y يتصف بصفة مميزة أو يقع ضمن فئة معينة $y_i = 1$

إذا كان خلاف ذلك $y_i = 0$

إذاً

$$Y = \sum_{i=1}^{N} y_i = A \quad \bar{Y} = \frac{1}{N}\sum_{i=1}^{N} y_i = \frac{A}{N} = \pi$$

حيث A تمثل عدد الوحدات في المجتمع التي تتصف بصفة مميزة و π تمثل نسبة الوحدات للذين يحملون صفة مميزة أو ينتمون إلى فئة معينة في المجتمع كذلك

$$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i-\pi)^2 = \frac{N\pi(1-\pi)}{N-1}$$

الآن نحن بصدد تقدير π واضح أن نسبة الذين يحملون صفة معينة أو ينتمون إلى فئة معينة في العينة تكون تقديراً إلى π أي

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{a}{n} = p$$

حيث إن a تمثل عدد الوحدات في العينة للذين يحملون صفة مميزة أو ينتمون إلى فئة معينة. إن برهنة p كتقدير غير متحيز إلى π ، أي $E(p)=\pi$ ليس بالصعب، نستطيع أن نستخدم نفس الطريقة التي استخدمناها لبرهنة أن الوسط الحسابي للعينة $\bar{y}$ تقدير غير متحيز إلى الوسط الحسابي للمجتمع $\bar{Y}$ أو باستخدام الطرق الرياضية راجع Cochran (1977) أو ترجمة أنيس كنجو (1995). أما تباين p فهو

$$V(p)=\frac{1-f}{n}S^2=\frac{N-n}{N-1}\,\frac{\pi(1-\pi)}{n}$$

ولكن π غير معلومة لذلك فإن S^2 غير معلومة. نستطيع تقدير S^2 باستخدام تباين العينة:

$$s^2=\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2=\frac{np(1-p)}{n-1}$$

لذلك فإن تقدير تباين p سيكون:

$$s^2(p)=\frac{1-f}{n-1}p(1-p)$$

إذا أهملنا معامل التصحيح فإن $s^2(p)$ يصبح:

$$s^2(p)=\frac{p(1-p)}{n-1}$$

أما تقدير الخطأ المعياري لتقدير p

$$s(p)=\sqrt{\frac{p(1-p)}{n-1}}$$

بقي أن نجد فترة ثقة للنسبة π إذا كان $n\pi \geq 5$ و $n(1-\pi)\geq 5$ فإن p يتبع التوزيع الطبيعي التقريبي أي أن:

$$p \cdot N\left(\pi,\ \frac{1-f}{n}\pi(1-\pi)\right)$$

إن فترة ثقة تقريبية للنسبة π فستكون:

$$p \mp z_{\alpha/2}\sqrt{\frac{1-f}{n-1}p(1-p)}$$

10.5 اختيار حجم العينة

يُعد اختيار حجم العينة المناسب مهماً جداً في النتائج النهائية للدراسة التي نقوم بها، حيث إنه كلما كبر حجم العينة كانت التقديرات والنتائج التي نحصل عليها دقيقة وقريبة من معلمات المجتمع. ولكن في المقابل لا نستطيع أن نأخذ حجم العينة كبيراً؛ لأن ذلك يتطلب جهداً ووقتاً وتكاليف مادية كبيرة قد لا يكون بالمستطاع توفيرها لحجم عينة كبير. فإذاً السؤال كيف نختار حجم عينة – وأقصد بالعينة هنا العينة العشوائية البسيطة– من المجتمع لتأمين دقة معينة للنتائج، وفي الوقت نفسه يكون بالإمكان تنفيذها على أرض الواقع وضمن الإمكانيات المتاحة؟ سوف يقتصر كلامنا على اختيار حجم العينة في ثلاث حالات نتناولها بالتفصيل فيما يأتي

1.10.5 اختيار حجم العينة لتقدير الوسط الحسابي للمجتمع

إن زيادة حجم العينة سيؤدي إلى زيادة الدقة للوسط الحسابي للعينة □ كتقدير للوسط الحسابي للمجتمع $\overline{Y}$ ولكن هذه الزيادة لها حدود في بعض الأحيان لا نستطيع تجاوزها، إن حجم عينة كبيرة يعني إضاعة للموارد المتاحة. كذلك حجم عينة صغيرة على الأرجح سوف يعطينا تقديراً غير دقيق وعلى هذا فمن الناحية المثالية باستطاعتنا أن نحدد الدقة التي نريدها أو الحد الأعلى للتكاليف التي نود صرفها، ومن ثم نحدد حجم العينة المطلوبة لتحقيق هذين الشرطين.

لكن هذا لن يتم بسهولة، حيث إن هناك كثيراً من المشكلات فمثلاً كيف نحدد الدقة التي نحتاجها للتقديرات المختلفة؟ كيف نوازن بين صفات أو متغيرات المجتمع المختلفة التي نريد أن ندرسها؟ كيف نتعامل مع بعض معلمات المجتمع غير المعروفة والتي نحتاجها –مثل تباين المجتمع– والتي ربما تؤثر في الدقة للتقديرات.

سوف تدرس حالة سهلة ولنفرض أننا نهدف لتقدير الوسط الحسابي للمجتمع لمتغير واحد وهو y باستخدام الوسط الحسابي للعينة العشوائية البسيطة. لذلك لا بد من تحديد الخطأ في تقدير الوسط الحسابي للمجتمع $\overline{Y}$ بمقدار معين وليكن d بحيث

$$P\left(\left|\overline{Y}-\overline{y}\right|\leq d\right)=1-\alpha$$

حيث إن d رقم محدد سلفاً و (α) صغير و $(1-\alpha)$ تسمى درجة الثقة. وباستخدام التوزيع الطبيعي التقريبي نستطيع أن نحدد الحد الأدنى لحجم العينة لتقدير الوسط الحسابي للمجتمع بما يأتي:

$$n\geq\frac{n_0}{1+n_0/N}$$

حيث إن

$$n_0=\frac{z_{\alpha/2}^2S^2}{d^2}$$

لكن S^2 تكون في الغالب غير معلومة وهناك أربعة طرق نستطيع أن نقدر بها S^2.

اختبار الموقع (Pilot Studies)

في أغلب الأحيان هذه الدراسة تجري قبل بدء عملية المسح الشامل أو المسح بالعينة، لتخدم أغراضاً متعددة، منها دراسة عدة إطارات للمعاينة إذا

كانت هذه الدراسة تأخذ شكل عينة عشوائية بسيطة يمكن استخدام نتائجها لكي تعطينا فكرة عن قيمة S^2. ولكن إذا كانت هذه الدراسة لا تأخذ بمبدأ العينة الاحتمالية فإذاً يجب التحفظ على استخدام نتائجها لمعرفة أو إعطاء فكرة عن S^2.

أ. استخدام مسوحات سابقة

ليس من غير الشائع أن نجد مسوحات سابقة قامت بها جهات معينة لدراسة المجتمع نفسه لصفات أو متغيرات مماثلة، وهذا يحدث كثيراً في ميدان التربية والطب وعلم النفس. في أغلب الأحيان يمكن استخدام البيانات من هذه الدراسات لتقدير S^2 .

ب. استخدام عينة عشوائية أولية

إن هذه الطريقة تُعَد من أفضل الطرق لتقدير S^2 ، ولكنها قد لا تكون ممكنة من الناحية الإدارية أو من ناحية التكاليف، ويمكن تلخيصها بأن تؤخذ عينة عشوائية بسيطة أولية بحجم n_1 أي بحجم كافٍ للحصول على تقدير مقبول من ناحية الدقة لتباين المجتمع S^2. وبعد تحديد حجم العينة الكلي n بالاعتماد على S_1^2 الذي وجد باستخدام العينة الأولية بحجم n_1. نقوم بسحب عينة عشوائية بسيطة بحجم $(n\square n_1)$ ، أي أن الوحدات التي حصلنا عليها باستخدام العينة الأولية سوف تدخل ضمن العينة العشوائية البسيطة النهائية. إن هذه الطريقة من أفضل الطرق، وتعد طريقة موضوعية وممكنة التطبيق، وكذلك يمكن الاعتماد على النتائج التي تم الحصول عليها باستخدام هذه الطريقة.

ت. استخدام بعض المعلومات عن ترتيب المجتمع

في بعض الأحيان نعرف معلومات عن المجتمع تحت الدراسة، يمكن استخدام هذه المعلومات لإعطائنا فكرة عن S^2 فمثلاً:

1. عدد الأخطاء المرتكبة في الصفحة الواحدة للكتب ذات الأحجام المتساوية والتي طبعت من قبل دار نشر معينة خلال مدة معينة. أو
2. عدد الأعطال التي تحدث في مسجل صوت من نوع معين خلال السنة الأولى من استعماله.

في كلتا الحالتين نستطيع أن نقول إنهما تتبعان توزيع بواسون، ومن معلوماتنا عن توزيع بواسون فإن $\overline{Y} = S^2$ يمكن استخدام هذه الخاصية، كذلك المعلومات عن متغيرات مشابهة من مسوحات سابقة نستطيع أن نعرف أو نقدر S^2 .

2105 اختيار حجم العينة لتقدير المجموع الكلي للمجتمع

لنفترض أننا نريد أن نقدر حجم العينة n لتقدير مجموع المجتمع الكلي Y باستخدام Y من العينة العشوائية البسيطة بحيث

$$P(\left| Y - Y \right| > d_1) < \alpha$$

حيث إن d_1 هنا تمثل مقدار حد الخطأ لتقدير Y. وباستخدام التوزيع الطبيعي التقريبي نحصل

$$n \geq \frac{n_0}{1 + n_0/N}$$

حيث إن:

$$n_0 = \frac{N^2 z^2_{\alpha/2} S^2}{d_1^2}$$

بعض الأحيان يعبر عن الدقة لتقدير $\overline{Y}$ أو Y على شكل نسبة. فعلى سبيل المثال يمكن أن نسأل ما هو حجم العينة بحيث إن:

$$P(\left| Y - Y \right| > \pi Y) \leq \alpha$$

فمثلاً ممكن أن نريد حجم العينة بحيث نكون متأكدين 95% أن Y سيكون ضمن 2% للمجموع الكلي الحقيقي للمجتمع Y هنا $\pi = 2\%$ و $\alpha = 0.05$ لابد من استبدال d_1 بما يساويها وهو πY إذا كان المتغير y تقريبا يتبع توزيع بواسون فإن:

$$S^2 = \overline{Y} = \frac{Y}{N}$$

وسيكون حجم العينة المطلوب

$$n \ge N\left[1 + (\pi / z_{\alpha/2})^2 Y\right]^{-1}$$

مثال (2): لمعرفة مقدار مبيعات سلسلة من المحلات التجارية مكونة من 243 محلاً. ما هو حجم العينة الذي نحتاجه لتقدير مجموع المبيعات ليكون ضمن 10% من الرقم الحقيقي مع ثقة أو دقة 95٪، من معلوماتنا عن مبيعات السنوات الثلاث الماضية عن مجموع المبيعات والانحراف المعياري للمبيعات بآلاف الريالات هي:

السنة	1	2	3
Y	321.7	366.8	401.0
S	0.826	0.776	0.804

$$n \ge 243\left[1 + \frac{1}{243}\left(\frac{0.10(363.17)}{0.8(1.96)}\right)^2\right]^{-1} = 76$$

3.10.5 اختيار حجم العينة لتقدير النسبة

بإمكاننا استخدام الطريقة التي استخدمناها في حالة تقدير الوسط الحسابي $\overline{Y}$ مع أخذ قيمة $S^2 = \pi(1 - \pi)$ لتحديد أو تقدير حجم العينة المطلوب لتقدير النسبة π وإذا أمعنا النظر في تباين تقدير النسبة p

$$V(p) = \frac{1-f}{n} S^2 = \frac{N-n}{N-1} \frac{\pi(1-\pi)}{n}$$

سنجد ان أعلى قيمة لها عندما تكون $\pi = (1-\pi) = 1/2$ ذلك عندما نعطي قيمة محددة لحجم العينة n سوف نقدر π بأقل دقة عندما تكون قيمته قريباً من 0.5، إذا أهملنا معامل التصحيح فإن حجم العينة في حده الأدنى يكون:

$$n_0 \geq \frac{z_{\alpha/2}^2 \pi(1-\pi)}{d^2}$$

وخلاف ذلك فإن حجم العينة يكون:

$$n \geq \frac{n_0}{1+n_0/N}$$

كذلك سوف نحصل على أكبر قيمة ممكنة لحجم العينة عندما تكون $\pi = (1-\pi) = 1/2$ في حالة تقدير النسبة فإن حجم العينة التي سنختارها عندما تكون قيمة $\pi = 0.5$ سيكون هو الحد الأعلى لحجم العينة المطلوب.

إذا كانت قيمة π غير معروفة في هذه الحالة يمكن وفي أسوأ الحالات أن نأخذ $\pi = 0.5$ أو نحاول معرفة قيمة π بالطرق نفسها التي اتبعناها لمعرفة قيمة S^2.

مثال (3): أوجد قيمة n بحيث تكون قيمة ضمن $\pm 2\%$ وثقه 95%.

إذا فرضنا $\pi = 0.5$ فإن

$$n_0 \geq \frac{(1.96)^2(0.5)(0.5)}{(0.02)^2} = 2401$$

ولكن إذا كانت $\pi = 0.1$ فإن

$$n_0 \geq \frac{(1.96)^2(0.1)(0.9)}{(0.02)^2} = 865$$

تمارين

1. إذا كان لدينا مجتمع يتكون من الوحدات A, B, C, D, E, F وهذه الوحدات تمثل أشخاصاً مصروفاتهم الشهرية بالريال كما يأتي:

الوحدة	A	B	C	D	E	F
المصروف الشهري	1200	1150	1400	1300	2000	1700

أ. أثبت باستخدام جميع العينات العشوائية البسيطة وبحجم 3 وحدات أن الوسط الحسابي للعينة وتباين العينة تقدير غير متحيز للوسط الحسابي ولتباين المجتمع أي أن $E(\bar{y}) = \bar{Y}$ و $E(s^2) = S^2$.

ب. إذا أردنا تقدير نسبة الأشخاص الذين مصروفاتهم الشهرية أكثر من 1400 ريال أثبت أن نسبة العينة التي هي تقدير للنسبة الحقيقية للمجتمع أنها تقدير غير متحيز أي $E(p) = \pi$.

2. سحبت عينة عشوائية بسيطة بحجم وحدتين من مجتمع متكون من $N=5$، قيم المتغير y_i هي 1, 0, □1, 5. احسب الوسط الحسابي $\bar{y}$ لجميع العينات الممكنة السحب، وأثبت أن تقدير الوسط الحسابي للعينة هو تقدير غير متحيز إلى الوسط الحسابي للمجتمع $\bar{Y}$.

3. للمجتمع في السؤال السابق احسب تباين العينة s^2 لجميع العينات الممكنة السحب وأثبت أنه تقدير غير متحيز لتباين المجتمع S^2.

4. إذا كان لدينا البيانات الآتية التي تمثل علامات $N=100$ طالب دخلوا الامتحان الأول لمساق مبادئ الإحصاء:

5, 6, 7, 10, 2, 7, 3, 7, 4, 5, 7, 4, 9, 6, 0, 3, 7, 2, 9, 3, 6, 8, 8, 4, 9, 10, 6, 3, 7, 5, 7 6, 8, 7, 5, 2, 4, 8, 4, 7, 4, 4, 7, 7, 4, 5, 8, 5, 4, 6, 7, 10, 4, 5, 7, 4, 1, 7, 9, 8, 6, 5, 8, 2, 2, 4, 7, 5, 10, 4, 6, 8, 4, 7, 8, 9, 3, 10, 7, 4, 7, 5, 3, 6, 6, 4, 9, 7, 7, 10, 5, 0, 1, 4, 5, 7, 8, 3, 2, 6

أ. أوجد معدل علامات الطلبة $\bar{Y}$ وتباين S^2 العلامات .

ب. اسحب عينة عشوائية بسيطة حجمها 5 طلاب، ثم أوجد الوسط الحسابي للعينة $\bar{y}$ وتباين. s^2 العينة وتباين $\bar{y}$ وتقدير تباين $\bar{y}$. قارن بين النتائج التي حصلت عليها باستخدام العينة العشوائية البسيطة والنتائج التي حصلت عليها في الفرع أ.

ت. كرر العملية التي قمت بها في الفرع ب وذلك باستخدام أحجام عينات 10، 20، 40 ثم قارن بين هذه النتائج ونتائج الفرع أ.

ث. اسحب 25 عينة عشوائية بحجم 5 وحدات لكل عينة، ثم أوجد الوسط الحسابي لأوساط العينات والتباين لأوساط العينات. قارن بين النتائج التي سنحصل عليها هنا ونتائجك في الفرع أ.

5. سحبنا عينة عشوائية بسيطة حجمها 12 قطعة أرض زراعية مزروعة بمحصول البطاطا، إذا كانت مساحة كل قطعة 1/4 دونم، قدر إنتاج الأرض الزراعية والتي مساحتها 500 دونم، ثم أوجد فترة 95% ثقة للمجموع الكلي للإنتاج.

4.1 , 4.7, 5.2, 3.9, 4.8, 3.8, 4.4, 6.0,5.3, 5.1 , 4.9, 3.3

6. سحبنا عينة عشوائية بسيطة حجمها 100 وحدة (مقياس الماء) في منطقة سكنية لتقدير معدل المصروفات الأسبوعية من الماء للعائلة، فحصلنا على $\bar{y} = 5.7$ متر مكعب و $s^2 = 425$. إذا كانت $N=10.000$ عائلة تسكن في المنطقة. أوجد فترة 90% ثقة للوسط الحسابي للمجتمع $\bar{Y}$. كذلك قدر المجموع الكلي لمصروفات المنطقة الأسبوعية من المياه، وأوجد معامل التغير لتقدير مجموع المياه المصروفة أسبوعياً بالمترات المكعبة.

7. تم سحب عينة عشوائية بحجم 13 عائلة من مدينة فيها 14.000 عائلة، عدد الأشخاص لكل عائلة في العينة هو: 4, 6, 5, 2, 7, 6, 3, 8, 10, 9, 3, 6, 4.

قدر عدد سكان المدينة، ثم أوجد 95% فترة ثقة لعدد سكان المدينة الحقيقي.

8. لمعرفة آراء طلبة إحدى الجامعات بنظام العلامات المستخدم في الجامعة، سحبت عينة عشوائية بسيطة حجمها 120 طالباً فوجدنا أن 80 طالباً ضد نظام العلامات المستخدم، قدر نسبة الطلبة المعارضين، وأوجد فترة 95% ثقة للنسبة الحقيقية π وكذلك أوجد معامل التغير للنسبة.

9. إذا اعتبرنا البيانات الواردة في السؤال الخامس تمثل عينة أولية من المجتمع، قدر حجم العينة المطلوبة لتقدير معدل إنتاج قطعة الأرض، إذا كان حد الخطأ للتقدير $d = \pm 1.5$ ودرجة الثقة 90%.

10. ما هو حجم العينة الذي يؤمن لنا حد خطأ للتقدير مقداره $d = \pm 0.03$ ودرجة ثقة 99% في السؤال الثامن أعلاه.

11. لتقدير إنتاج أحد بساتين البرتقال الذي يوجد فيه $N= 1500$ شجرة، قام المزارع بسحب عينة عشوائية بسيطة حجمها $n=10$ أشجار فوجد البيانات الآتية $\bar{y} = 40.5$ كغم و$s^2 = 40$. قَدِّر الإنتاج الكلي للبستان، ثم أوجد فترة 95% ثقة للمجموع الكلي.

12. إذا اعتبرنا أن البيانات في السؤال الحادي عشر تمثل عينة أولية، أوجِدْ حجم العينة لتقدير مجموع الإنتاج الكلي إذا كان حد الخطأ في تقدير المجموع $d=1000$.

13. يبلغ عدد سكان إحدى المدن الصغيرة 5000 شخص. ترغب دائرة الصحة في المدينة أن تقدر نسبة الأشخاص في الفرية الذين تكون فصيلة دمهم AB. ما هو حجم العينة التي يمكن سحبها إذا كنا نرغب بأن يكون تقدير النسبة الحقيقية بخطأ مقداره ±4% وبدرجة ثقة مقدارها 99%. علماً بأننا سحبنا عينة عشوائية أولية فوجدنا نسبة الذين يحملون فصيلة الدم AB حوالي 20%.

14. سحبت عينة عشوائية حجمها 5 طلاب من كشف يحتوي على 50 طالباً مسجلين في مادة الإحصاء، الاستبانة التي سلمت لهم تحتوي على سؤالين هما:

السؤال الأول: هل تمتلك سيارة؟ نعم □□□□□□ لا □□□□□□

السؤال الثاني: ما هو مصروفك بالدولار ألأمريكي؟ □□□□□□□ دولار.

يعطي الجدول الآتي إجابات الطلبة

الطالب	السؤال الأول	السؤال الثاني
1	لا	300
2	نعم	1500
3	لا	1000
4	لا	500
5	نعم	1200

أ. قدر معدل ومجموع مصروف الطلاب الشهري. ثم أوجد 95% فترة ثقة لمعدل ومجموع مصروفاتهم الحقيقية لكل شهر.

ب. أوجد الخطأ المعياري لتقدير نسبة الطلاب الذين يمتلكون سيارة ثم أوجد 93% فترة ثقة للنسبة الحقيقية للطلاب الذين يمتلكون سيارة.

15. قدر مصروفات الطلبة الشهرية لطلبة إحدى كليات الجامعة.

16. إذا طلبت منك إدارة المطاعم الموجودة في الجامعة القيام باستطلاع آراء الطلبة الذين يتناولون وجبات غذائية في أحد مطاعم الجامعة حول آرائهم في الوجبات الغذائية المقدمة في هذا المطعم. صمم استمارة استبانه ثم نفذها وذلك بجمع البيانات اللازمة لمعرفة آراء الطلبة حول الوجبات الغذائية المقدمة لهم.

17. ترغب غرفة تجارة الدمام معرفة آراء أعضائها عن الخدمات التي تقدمها الغرفة إلى أعضائها. إذا طلب منك تنفيذ هذا المشروع، ما الخطوات اللازمة لإنجازه ابتداء من تحديد الأهداف إلى تحليل البيانات؟
18. حدد إحدى المشكلات في مجال عملك أو دراستك، بحيث تستطيع أن تسحب عينة عشوائية بسيطة لتقدير الوسط الحسابي للمجتمع والمجموع الكلي أو النسبة.

المراجع العربية

1. سليمان محمد طشطوش (2001)، أساسيات المعاينة الإحصائية، دار الشروق للنشر والتوزيع، عمان.
2. حسين علوان (1994)، طرق المعاينة، دار الفرقان، عمان.
3. هلال عبود البياتي وصبري رديف العاني (1981)، العينات، جامعة بغداد، بغداد.
4. وليد عبدالحميد نوري وعبدالمجيد حمزة الناصر (1981) العينات، وزارة التعليم العالي والبحث العلمي، بغداد.
5. وليم كوكوران –ترجمة أنيس كنجو– تقنية المعاينة الإحصائية (1995)، جامعة الملك سعود، الرياض.

References

1. Barnett, V. (1991). Sampling Survey Principles and Methods, Edward Arnold, London.
2. Benedetto, J. J. and Ferreira, P. J. (2001). Modern Sampling Theory, Birkhauser Boston.
3. Baus, D. (1958). On Sampling with and without Replacement, Sankhya 20, p 287□294.
4. Chaudhuri, A. and Stenger, H. (1992). Survey Sampling□Theory and Methods, Marcel Dekker, New York.
5. Cochran, W. G. (1977). Sampling Technique, 3rd Ed. Wiley, New York.
6. Deming, W. E. (1960). Sampling Design in Business Research. Wiley, New York.
7. Foreman, E. K. (1991). Survey Sampling Principles, Marcel Dekker, New York.
8. Govindarajulu, Z. (1999). Elements of Sampling Theory and Methods, Prentice Hall.
9. Hajek, J (1960). Limiting Distributions in Simple Random Sampling from Finite Population. *Pub. Math. Inst. Hungarian Acad. Sci.* 5, 361□374.
10. Hajek, J (1981). Sampling from Finite Population, Marcel Dekker, New York.
11. Hansen, M.H., and Hurwitz, W. N. (1943). On the theory of sampling from finite popula□ions. Ann. Math. Statist. 14,333□362.
12. Hedayat, A. S. and Sinha, B. K. (1991). Design and Inference in Finite Population Sampling, Wiley, New York.
13. Kish, L (1995). Survey Sampling, Wiley, New York.
14. Lohr, S. L. (1999). Sampling□Design and Analysis, Duxbury, New York.
15. Levy, P. S. and Lemeshow S (1991) Sampling of Population□Methods and Applications, Wiley, New York.
16. Mahalanobis, P. C. (1946). Recent experiments in statistical sampling in the Indian Sta□tistical Institute. J Roy. Statist. Soc. A109, 325□378, reprinted in Sankhyii.
17. Murthy, M. N. (1962). Variance and confidence interval estimation Sankhyii 24(B), 1□12.
18. Raj, D (1968). Sampling Theory, McGraw□Hill, New York.
19. Rao, J. N. K. (1985). Conditional inference in survey sampling. Survey Methodology, D,I. Inst. Statist. Math. No.1, 15□31.
20. Sampath, S. (2000). Sampling Theory and Methods, CRC Press.

21. Scheaffer, R.L., Mendenhall, W. and Ott L. (1996). Elementary Sampling Survey, 5^{th} ed., Duxbury, New York.
22. Smith, T. M. F. (1976). The Foundation of Survey Sampling□A review, *Journal of Royal Statistics Society* A139, p183□204
23. Sturat, A. (1984). The Ideas of Sampling, Revised Edition, Griffin, London.
24. Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., and Ashok, C. (1984). Samplmg Theory of Surveys with Applications, 3d ed., Ames (Iowa)□Iowa State University Press.
25. Thompson, S. K. (2002). Sampling, 2end Wiley, New York.
26. Tryfos, P. (1996). Sampling Methods for Applied Research, Wiley, New York.
27. Yates, F. (1981). Sampling Methods for Censuses and Surveys 4th Ed., Grittin, London.

الفصل السادس

العينة العشوائية الطبقية
Stratified Random Sample

1.6 مقدمة

إن المسح بطريقة العينة العشوائية الطبقية يُعد من أكثر طرق المعاينة استخداماً. حيث يُقسم المجتمع في المعاينة العشوائية الطبقية الذي يتكون من N من الوحدات إلى K من المجتمعات الجزئية تدعى طبقات. إن المجتمع الجزئي i يحتوي على N_i من الوحدات $i=1,2,\dots,K$ يجب أن تكون المجتمعات الجزئية غير متداخلة حتى يكونوا المجتمع أي:

$$N = N_1 + N_2 + \dots + N_k$$

تُعرف العينة العشوائية الطبقية على أنها: العينة التي اختيرت عند تقسيم عناصر المجتمع إلى مجتمعات جزئية وغير متداخلة تدعى طبقات، وبعد ذلك نختار عينة عشوائية بسيطة من كل طبقة. نختار العينة من كل طبقة بصورة مستقلة، حجم العينة الذي يختار من الطبقة i هو n_i و $i=1,2,\dots,K$ بحيث $n = n_1 + n_2 + \dots + n_k$. إذا العينة اختيرت من كل طبقة بشكل عشوائي عندئذ تسمى العينة العشوائية الطبقية.

إن الهدف الرئيس من الطبقية ليعطينا فهم ودراية أفضل بأجزاء مختلفة من المجتمع ومن ثم نحصل على دقة أعلى. للوصول إلى تحقيق الهدف يجب أن

يقسم المجتمع إلى طبقات، ويحدد عدد الطبقات، وكذلك توزع العينة بين الطبقات، وأخيراً تحلل البيانات على أنها من عينة طبقية.

2.6 أسس تقسيم المجتمع إلى طبقات

يمكن تلخيص الأسس التي تتبع لتقسيم المجتمع إلى طبقات بما يأتي:

1. يجب أن تكون الطبقات غير متداخلة، ويجب أن تؤلف مع بعضها المجتمع كاملاً.
2. يجب أن تتم عملية تقسيم المجتمع إلى طبقات بحيث إن كل طبقة تكون متجانسة بالنسبة إلى الصفة التي تجري دراستها.
3. عندما تصعب عملية تقسيم المجتمع إلى طبقات بناء على الصفة التي تجري دراستها يكون تقسيم المجتمع إلى طبقات على أساس ملاءمة التقسيم بالنسبة إلى النواحي الإدارية.
4. إذا أعطيت حدود الدقة إلى بعض المجتمعات الجزئية من الأفضل أن تعامل هذه المجتمعات الجزئية على أنها طبقات.

3.6 مزايا تقسيم المجتمع إلى طبقات

إن تقسيم المجتمع إلى طبقات له مزايا عديدة أهمها:

1. ربما يكون تقسيم المجتمع إلى طبقات مفضلاً ذلك لأنه ملائم من النواحي الإدارية. حيث إن الجهة التي تقوم بالمسح يمكن أن تجد عاملين في مختلف المناطق الإدارية، وبالتالي يؤدي إلى تنظيم ومراقبة ومتابعة أفضل. وبالتالي ربما سيؤدي إلى تقليل تكاليف مشاهدة الوحدة الواحدة.
2. تقسيم المجتمع بناء على الصفات الطبيعية يساعد على تحسين تصميم المعاينة. فمثلاً في حالة مسح الحقول الزراعية ربما تكون هنالك

مشـكلات معاينـة في منـاطق السـهول تختلـف عمـا هـو موجـود في الصـحراء أو المنـاطق الجبليـة. لـذلك يكـون مـن الأفضـل أن يقسـم المجتمع إلى طبقات بحسب المناطق الجغرافية المختلفة.

3. إذا كـان المجتمـع مكونـاً مـن عـدد كـبير مـن العناصـر سـوف يـؤدي تقسيمه إلى طبقات إلى تقليل التشتت داخل الطبقـات، ومـن ثـم سـوف نحصل على تقديرات أكثر دقة باستخدام العينـة الطبقيـة، وذلـك مـن خلال القيام بالتقديرات داخل كل طبقة، ومـن ثـم جمعهـا مـع بعضـها البعض للوصول إلى التقدير النهائي.

4. إن تقسـيم المجتمـع إلى طبقـات يعطـي مجموعـات مختلفـة مـن المجتمـع فرصة للظهور في العينة والـتي قـد تـكون مهمـة. ويسـاعد علـى اختيـار أفضل لأجزاء مختلفة للمجتمع.

5. نحصل على دقة أكبر للتقديرات المختلفـة لمعلمـات المجتمـع، وذلـك لأن المجتمع يكون غير متجانس بصورة عامـة، ولكـن عنـدما يقسـم إلى أجـزاء (طبقـات) متجانسـة مـع بعضـها، هـذا يـؤدي إلى الحصـول علـى تقديرات أكثر دقة داخل كل طبقة ومن ثم إلى المجتمع كله.

4.6 رموز ومصطلحات

لنفترض أن i يرمـز إلى الطبقـة وj يرمـز إلى وحـدات المعاينـة في الطبقـة i، المصطلحات الآتية تعود إلى الطبقة i

N_i : عدد الوحدات في الطبقة

n_i : عدد الوحدات في العينة المسحوبة من الطبقة

$W_i=N_i/N$: وزن الطبقة

f_i : كسر المعاينة في الطبقة

لنفترض أن y_{ij} يمثل القيمة للوحدة j الموجودة في الطبقة i

$Y_i = \sum_{j=1}^{N_i} y_{ij}$: المجموع الكلي للطبقة

$\bar{Y}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} y_{ij}$: الوسط الحسابي للطبقة

$\bar{y}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} y_{ij}$: الوسط الحسابي للعينة المسحوبة من الطبقة

$Y_i = N_i \bar{y}_i$: تقدير المجموع الكلي في الطبقة

$S_i^2 = \frac{1}{N_i - 1}\sum_{j=1}^{N_i}(y_{ij} - \bar{Y}_i)^2$: تباين الطبقة

$s_i^2 = \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(y_{ij} - \bar{y}_i)^2$: تباين العينة المسحوبة من الطبقة

$var(\bar{y}_i) = (1 - f_i)\frac{S_i^2}{n_i}$: تباين الوسط الحسابي للعينة المسحوبة من الطبقة

$var(Y_i) = \frac{N_i^2}{n_i}(1 - f_i) S_i^2$: تباين تقدير المجموع الكلي للطبقة

$Y = \sum_{i=1}^{K}\sum_{j=1}^{N_i} y_{ij}$: المجموع الكلي للمجتمع

$\bar{Y} = \frac{Y}{N}$: الوسط الحسابي للمجتمع

5.6 تقدير الوسط الحسابي والمجموع الكلي للمجتمع

1.5.6 تقدير الوسط الحسابي للمجتمع

إن الوسط الحسابي للعينة الطبقية $\bar{y}_{st}$ هو التقدير البدهي للوسط الحسابي للمجتمع $\bar{Y}$ حيث يعرف

$$\bar{y}_{st} = \sum_{i=1}^{K}\frac{N_i}{N}\bar{y}_i = \sum_{i=1}^{K} W_i \bar{y}_i$$

إنه من السهولة ملاحظة أن $\bar{y}_{st}$ تقدير غير متحيز إلى $\bar{Y}$ يراجع Cochran(1977) أو ترجمة أنيس كنجو (1995) لهذا البرهان الرياضي

وكذلك جميع البراهين الرياضية الأخرى في هذا الفصل. أما تباين $\bar{y}_{st}$ فيمكن إيجاده بسهولة

$$Var(\bar{y}_{st}) = \sum_{i=1}^{K} W_i^2 (1 - f_i)\, S_i^2 / n_i = \sum_{i=1}^{K} \frac{W_i^2 S_i^2}{n_i} - \sum_{i=1}^{K} \frac{W_i S_i^2}{N}$$

إن إيجاد تباين الوسط الحسابي للعينة الطبقية يقتضي معرفة تباين الطبقات S_i^2 وهذا غير ممكن لأنه في الغالب S_i^2 غير معروفة؛ لذلك لا بد من تقدير تباين الوسط الحسابي للعينة الطبقية وذلك باستخدام s_i^2 تباين العينة المسحوبة من الطبقة بدلاً من S_i^2 تباين الطبقة، ليصبح تقدير تباين $\bar{y}_{st}$ كما يأتي:

$$s^2(\bar{y}_{st}) = \sum_{i=1}^{K} W_i^2 (1 - f_i)\, s_i^2 / n_i = \sum_{i=1}^{K} \frac{W_i^2 s_i^2}{n_i} - \sum_{i=1}^{K} \frac{W_i\, s_i^2}{N}$$

كذلك فإن تقدير الخطأ المعياري:

$$s(\bar{y}_{st}) = \sqrt{s^2(\bar{y}_{st})}$$

أما معامل التغير إلى $\bar{y}_{st}$ فيمكن إيجاده:

$$CV = s(\bar{y}_{st}) / \bar{y}_{st}$$

وأخيراً يمكن استخدام التوزيع الطبيعي التقريبي إذا كان حجم العينة كبيرًا لإيجاد $(1 - \alpha)100\%$ فترة ثقة للوسط الحسابي للمجتمع:

$$\bar{y}_{st} \pm z_{\alpha/2}\, s(\bar{y}_{st})$$

2.5.6 تقدير المجموع الكلي للمجتمع

لتقدير المجموع الكلي للمجتمع يمكن إيجاده باستخدام:

$$Y_{st} = N\bar{y}_{st} = \sum_{i=1}^{K} N_i\, \bar{y}_i$$

أما تباين Y_{st} فيمكن إيجاده

$$Var(Y_{st}) = \sum_{i=1}^{K} \frac{N_i^2}{n_i}(1-f_i)S_i^2$$

كذلك لا بد من تقدير تباين تقدير المجموع الكلي حيث (وكما مر معنا قبل قليل) إن S_i^2 غير معروفة ولا بد من تقديرها باستخدام s_i^2 ليصبح تقدير تباين Y_{st} كما يأتي

$$s^2(Y_{st}) = \sum_{i=1}^{K} \frac{N_i^2}{n_i}(1-f_i)s_i^2 = \sum_{i=1}^{K} \frac{N_i^2 s_i^2}{n_i} - \sum_{i=1}^{K} N_i s_i^2$$

مثال (1): لتطبيق المعادلات التي وردت في هذا الفصل، لنفترض أن لدينا الجدول الآتي:

Y_i	W_i	N_i	s_i^2	$\bar{y}_i$	n_i	الطبقة
35000	0.2	700	225	50	40	1
52500	0.5	1750	64	30	50	2
21000	0.3	1050	36	20	40	3

الحل:

تقدير المجموع الكلي

$$Y_{st} = \sum_{i=1}^{3} N_i \bar{y}_i = \sum_{i=1}^{3} Y_i = 108500$$

تقدير تباين تقدير المجموع الكلي

$$s^2(Y_{st}) = \sum_{i=1}^{K} \frac{N_i^2 s_i^2}{n_i} - \sum_{i=1}^{K} N_i s_i^2 = 7361200$$

تقدير الخطأ المعياري لتقدير المجموع الكلي

$$s(Y_{st}) = \sqrt{s^2(Y_{st})} = 2173.13$$

الوسط الحسابي للعينة الطبقية

$$\bar{y}_{st} = \sum_{i=1}^{3} W_i \, \bar{y}_i = \frac{Y_{st}}{N} = 31$$

تقدير تباين الوسط الحسابي للعينة الطبقية

$$s^2(\bar{y}_{st}) = \sum_{i=1}^{K} \frac{W_i^2 s_i^2}{n_i} - \sum_{i=1}^{K} \frac{W_i^2 s_i^2}{N} = 0.6009$$

تقدير الخطأ المعياري

$$s(\bar{y}_{st}) = \sqrt{s^2(\bar{y}_{st})} = 0.78$$

تقدير معامل التغير إلى $\bar{y}_{st}$

$$CV = \frac{s(\bar{y}_{st})}{\bar{y}_{st}} = \frac{0.78}{31} = 0.025$$

فترة 95% ثقة للوسط الحسابي $\bar{Y}$

$$\bar{y}_{st} \pm z_{\alpha/2} s(\bar{y}_{st}) = 31 \pm 1.96(0.78) \square (29.47, 32.53)$$

6.6 توزيع حجم العينة بين الطبقات

هنالك عدة طرق لتوزيع حجم العينة n بين الطبقات، ولكن سوف نتناول أهم طريقتين. المعادلات التي أعطيت سابقاً تعتمد على قيمة n_i ، بحيث إن

$$n = \sum_{i=1}^{K} n_i$$

1.6.6 التوزيع المتناسب (Proportional Allocation)

إن فكرة التوزيع المتناسب تقوم على أساس توزيع حجم العينة n بين الطبقات بحيث يكون نصيب كل طبقة من n متناسباً مع حجم الطبقة في

المجتمع، أي أن $n_i = nW_i$ وإذا عوضنا قيمة في المعادلة الخاصة بتباين الوسط الحسابي للعينة الطبقية نحصل على

$$V_{prop}(\bar{y}_{st}) = \sum_{i=1}^{K} \frac{W_i S_i^2}{n} - \frac{1}{N}\sum_{i=1}^{K} W_i S_i^2$$

وهذا يسمى تباين الوسط الحسابي للعينة الطبقية باستخدام التوزيع المتناسب.

2.6.6 التوزيع الأمثل (Optimum Allocation)

لإيجاد أفضل توزيع لحجم العينة n في الحقيقة هو عبارة عن تصغير Minimizing دالة التباين $Var(\bar{y}_{st})$ لحجم العينة مع بعض الشروط. ويمكن أن نفترضها دالة لحجم العينة n_i

$$Var(\bar{y}_{st}) = \sum_{i=1}^{K} \frac{W_i^2 S_i^2}{n_i} - \sum_{i=1}^{K} \frac{W_i S_i^2}{N}$$

للحصول على أقل تباين ممكن راجع Cochran (1977) أو ترجمة أنيس كنجو(1995) للبرهان وتفصيلات أخرى. إذا كانت تكاليف مشاهدة الوحدة الواحدة في جميع الطبقات متساوية فإن التوزيع الأمثل لحجم العينة بين الطبقات يكون

$$n_i = n \frac{W_i S_i}{\sum_{i=1}^{K} W_i S_i}$$

إذا عوضنا عن قيمة n_i في تباين الوسط الحسابي للعينة الطبقية نحصل على ما يسمى بالتباين الأمثل

$$V_{opt}(\bar{y}_{st}) = \frac{[\sum_{i=1}^{K} W_i S_i]^2}{n} - \sum_{i=1}^{K} \frac{W_i S_i^2}{N}$$

كذلك يمكن الحصول على أفضل (أصغر) n مع كون التباين ثابتاً ومساوياً إلى V

$$n = \frac{\left[\sum_{i=1}^{K} W_i S_i\right]^2}{V + \frac{1}{N}\sum_{i=1}^{K} W_i S_i^2}$$

مثال (2): لنعد إلى المثال (1) ونقارن بين قيمة التباين للحالات الثلاثة أي توزيع n الحقيقي كما وردت في المثال وباستخدام التوزيع المتناسب والتوزيع الأمثل

الطبقة	التوزيع الحقيقي	التوزيع الأمثل	التوزيع المتناسب
1	40	44	26
2	50	59	65
3	40	27	39
مجموع حجم العينة	130	130	130
التباين	0.6009	0.570	0.6503

أما إذا كانت تكاليف مشاهدة الوحدة في الطبقات المختلفة غير متساوية فلا بد من معرفة دالة التكاليف، وتكاليف مشاهدة الوحدة الواحدة في كل طبقة. إن دالة التكاليف الشائعة الاستعمال في العينة الطبقية هي:

$$C = C_0 + \sum_{i=1}^{K} c_i n_i$$

حيث إن C تمثل إجمالي الميزانية، و C_0 تكاليف ثابتة و c_i تكاليف مشاهدة الوحدة الواحدة في الطبقة i. وباستخدام أسلوب التصغير (Minimizing) لدالة

التباين باستخدام دالة التكاليف نحصل على التوزيع الأمثل لحجم العينة بين الطبقات كما يأتي:

$$n_i = n \frac{W_iS_i/\sqrt{c_i}}{\sum_{i=1}^{K} W_iS_i/\sqrt{c_i}}$$

أما إذا كانت ميزانية المسح للعينة ثابتة أي أن

$$C' = C - C_o = \sum_{i=1}^{K} c_i n_i$$

باستخدام هذه الميزانية لا بد من إيجاد حجم العينة الأمثل وهو

$$n = C' \frac{\sum_{i=1}^{K} W_iS_i/\sqrt{c_i}}{\sum_{i=1}^{K} W_iS_i\sqrt{c_i}}$$

إذا كان التباين ثابتاً ومساوياً إلى V نجد قيمة n التي تعطينا التباين المطلوب

$$n = \frac{[\sum_{i=1}^{K} W_iS_i\sqrt{c_i}][\sum_{i=1}^{K} W_iS_i/\sqrt{c_i}]}{V + \frac{1}{N}\sum_{i=1}^{K} W_iS_i^2}$$

مثال (3): لنعد إلى المثال (1) ولنفترض أن تكاليف مشاهدة الوحدة الواحدة في الطبقات هي

$c_1 = 1, c_2 = 4, c_3 = 9$ أوجد تقدير التباين المثالي باستخدام هذه العلاقة

$$n_i = n \frac{W_iS_i/\sqrt{c_i}}{\sum_{i=1}^{K} W_iS_i/\sqrt{c_i}}$$

وجدنا $n_1 = 70, n_2 = 46, n_3 = 14$. الآن نعوض عن قيمة n_i في المعادلة الآتية لنجد تقديرًا لقيمة التباين الأمثل مع كون تكاليف مشاهدة الوحدة مختلفة من طبقة إلى أخرى.

$$s^2(\bar{y}_{st}) = \sum_{i=1}^{K} \frac{W_i^2 s_i^2}{n_i} \square \sum_{i=1}^{K} \frac{W_i s_i^2}{N} = 0.6827$$

لابد من الإشارة في نهاية هذا الجزء إلى أن

$$Var(\bar{y}) \geq V_{prop}(\bar{y}_{st}) \geq V_{opt}(\bar{y}_{st})$$

لمزيد من المعلومات راجع Cochran (1977) أو ترجمة أنيس كنجو(1995).

7.6 تقدير نسبة المجتمع

تقدير النسبة للمجتمع يمكن إيجاده بسهولة باستخدام التعريف الآتي

$y_{ij} = 1$ إذا كان العنصر j الموجود في الطبقة i يحمل صفة مميزة

$y_{ij} = 0$ خلاف ذلك أي إن العنصر j الموجود في الطبقة i لا يحمل الصفة المميزة

سوف نستخدم P_i للتعبير عن النسبة في الطبقة i و P للتعبير عن النسبة في المجتمع وكذلك $\hat{p}_i = \frac{a_i}{n_i}$ للتعبير عن تقدير النسبة في الطبقة i ، حيث إن a_i يمثل مجموع الوحدات من الذين يحملون الصفة المميزة في العينة المسحوبة من الطبقة. نستطيع الآن أن نعرف تقدير النسبة في العينة العشوائية الطبقية بما يأتي

$$\hat{P}_{st} = \sum_{i=1}^{K} W_i \hat{P}_i = \sum_{i=1}^{K} \frac{N_i}{N} \frac{a_i}{n_i}$$

يمكن أن نبرهن بسهولة أن $\hat{p}_{st}$ تقدير غير متحيز إلى نسبة المجتمع P وبتباين

$$Var(\hat{p}_{st}) = \sum_{i=1}^{K} \frac{W_i^2 P_i Q_i}{n_i} \square \sum_{i=1}^{K} \frac{W_i P_i Q_i}{N}$$

حيث إن $Q_i = 1 - P_i$. ولكن P_i غير معلومة لذلك لابد من تقدير $Var(\hat{p}_{st})$ بما يأتي

$$s^2(\hat{p}_{st}) = \sum_{i=1}^{K} \frac{W_i^2 \hat{p}_i \hat{q}_i}{n_i} - \sum_{i=1}^{K} \frac{W_i \hat{p}_i \hat{q}_i}{N}$$

و معامل التغير لتقدير النسبة في العينة العشوائية الطبقية

$$CV(\hat{p}_{st}) = \frac{s(\hat{p}_{st})}{\hat{p}_{st}}$$

حيث إن

$$s(\hat{p}_{st}) = \sqrt{s^2(\hat{p}_{st})}$$

ويسمى تقدير الخطأ المعياري للعينة الطبقية.

ملاحظة: يمكن بسهولة تقدير مجموع التكرارات للعينة العشوائية الطبقية باستخدام

$$A_{st} = N\hat{p}_{st}$$

تباين مجموع التكرارات

$$Var(A_{st}) = N^2 \, Var(\hat{p}_{st})$$

تقدير تباين مجموع التكرارات

$$s^2(A_{st}) = N^2 s^2(\hat{p}_{st})$$

معامل التغير لمجموع التكرارات

$$CV(A_{st}) = \frac{s(A_{st})}{A_{st}}$$

مثال (4): استخدام الجدول الآتي لحساب النتائج التي وصلنا إليها فيما يخص تقدير النسبة ، ومجموع التكرارات

الطبقة	N_i	n_i	W_i	p_i
1	700	40	0.2	0.5
2	1750	50	0.5	0.8
3	1050	40	0.3	0.1
المجموع	3500	130		

الحل:

تقدير النسبة الطبقية

$$\hat{P}_{st} = \sum_{i=1}^{3} W_i \hat{P}_i = 0.53$$

تقدير تباين النسبة الطبقية

$$s^2(\hat{p}_{st}) = \sum_{i=1}^{K} \frac{W_i^2 \hat{p}_i \hat{q}_i}{n_i} - \sum_{i=1}^{K} \frac{W_i \hat{p}_i \hat{q}_i}{N} = 0.0012$$

تقدير الخطأ المعياري للنسبة الطبقية

$$s(\hat{p}_{st}) = \sqrt{s^2(\hat{p}_{st})} = 0.0351$$

معامل التغير للنسبة الطبقية

$$CV(\hat{p}_{st}) = \frac{s(\hat{p}_{st})}{\hat{p}_{st}} = \frac{0.0351}{0.531} = 0.066$$

تقدير مجموع التكرارات

$$A_{st} = N\hat{p}_{st} = 3500(0.53) = 1855$$

تقدير تباين مجموع التكرارات

$$s^2(A_{st}) = N^2 s^2(\hat{p}_{st}) = 15123.13$$

معامل التغير لمجموع التكرارات

$$CV(A_{st}) = \frac{s(A_{st})}{A_{st}} = 0.066$$

إذا استخدمنا التوزيع المتناسب لتوزيع n بين الطبقات أي $n_i = nW_i$ نحصل على التباين المتناسب

$$V_{prop}(\hat{p}_{st}) = \sum_{i=1}^{K} \frac{W_i P_i Q_i}{n} - \sum_{i=1}^{K} \frac{W_i P_i Q_i}{N}$$

إذا استخدمنا التوزيع الأمثل وكذلك

$$S_i^2 = \frac{N_i}{N_i - 1} P_i Q_i \approx P_i Q_i$$

وإذا كانت التكاليف متساوية فإن

$$n_i = n \frac{W_i \sqrt{P_i Q_i}}{\sum_{i=1}^{K} W_i \sqrt{P_i Q_i}}$$

أما إذا كانت التكاليف غير متساوية فإن

$$n_i = n \frac{W_i \sqrt{P_i Q_i} \,/\, \sqrt{c_i}}{\sum_{i=1}^{K} W_i \sqrt{P_i Q_i} \,/\, \sqrt{c_i}}$$

أخيراً إذا كانت التكاليف متساوية نحصل على التباين الأمثل إلى $\hat{p}_{st}$ كما يأتي

$$V_{opt}(\hat{p}_{st}) = \frac{1}{n} \sum_{i=1}^{K} [W_i \sqrt{P_i Q_i}]^2 - \sum_{i=1}^{K} \frac{W_i P_i Q_i}{N}$$

8.6 اختيار حجم العينة

إن مقدار المعلومات الموجودة في العينة يعتمد على حجم العينة n بما أن $Var(\bar{y}_{st})$ يتناقص عندما يزداد حجم العينة n دعنا نختار الطريقة التي يمكن

أن نختار فيها حجم عينة معين لإيجاد كمية ثابتة من المعلومات لتقدير معلمات المجتمع.

1.8.6 اختيار حجم العينة لتقدير الوسطي الحسابي والمجموع الكلي للمجتمع

لنفترض أننا حددنا أن التقدير $\bar{y}_{st}$ يجب أن يكون ضمن d وحدة من الوسط الحسابي للمجتمع $\bar{Y}$ باحتمال تقريبا 0.95، أي إن حد خطأ التقدير للوسط الحسابي الطبقي للعينة هو d ودرجة الثقة 0.95 بعبارة أخرى

$$2\sqrt{V(\bar{y}_{st})} = d$$

أو

$$V(\bar{y}_{st}) = d^2/4$$

الملاحظ أن المعادلة الأخيرة تعتمد على $S_1^2, S_2^2, \ldots, S_k^2$ وهذه غير معروفة، ولكن يمكن استبدال هذه القيم بـ $s_1^2, s_2^2, \ldots, s_k^2$، إذا كان حجم المجتمع N كبير، كذلك نحن وضعنا: $V(\bar{y}_{st}) = d^2/4$ في الحقيقة لا نستطيع حل هذه المعادلة لإيجاد قيمة n ما لم نعلم العلاقة بين n و $n_1, n_2, \ldots, n_k$ وطبعاً هناك عدة طرق لتوضيح العلاقة، منها ما يسمى بالتوزيع المتناسب أي إن $n_i = nW_i$. وباستخدام هذه المعادلة نستطيع حل المعادلة $V(\bar{y}_{st}) = d^2/4$. وبالطريقة نفسها نستطيع إيجاد قيمة n لتقدير المجموع الكلي Y وذلك بحل المعادلة $V(\bar{y}_{st}) = d^2/4N^2$.

الآن باستطاعتنا إيجاد القيمة التقريبية لحجم العينة n لتقدير $\bar{Y}$ أو Y مع حد لخطأ التقدير مقداره d

$$n \geq \frac{\sum_{i=1}^{K} N_i^2 S_i^2 / W_i}{N^2 D + \sum_{i=1}^{K} N_i S_i^2}$$

عندما نقدر $\overline{Y}$ فإن $D=d^2/4$ وعندما نقدر Y فإن $D=d^2/4N^2$

لا بد من معرفة $S_1^2, S_2^2, ..., S_k^2$ وإلا لا بد من تقديرها باستخدام الطرق نفسها التي تكلمنا عنها في الفصل الخامس. إحدى هذه الطرق نقوم فيها باستخدام تباين العينة الأولية المسحوبة من الطبقة i للحصول على S_i^2 و $i=1, 2, \ldots, k$.

مثال (5): في المثال (1) لنفترض أنه من التجارب السابقة فإن تباين المجتمع للطبقات الثلاثة $S_1^2=300, S_2^2=81, S_3^2=40$ ، نرغب بتقدير الوسط الحسابي للمجتمع باستخدام الوسط الحسابي للعينة الطبقية، ما هو حجم العينة الذي نحتاجه بحيث إن الحد في مقدار الخطأ للتقدير يساوي 2.

إن حجم العينة المطلوب

$$n \geq \frac{\sum_{i=1}^{K} N_i^2 S_i^2 / W_i}{N^2 D + \sum_{i=1}^{K} N_i S_i^2} \approx 109$$

2.8.6 اختيار حجم العينة لتقدير نسبة المجتمع

لتقدير نسبة المجتمع أولاً نحدد مقدار المعلومات التي نرغب بها، وذلك من خلال تحديد مقدار حد الخطأ في تقدير النسبة، ومن ثم نختار حجم العينة بناء على ذلك إن الطريقة التي تستخدم لإيجاد حجم العينة n عندما يحدد الحد في خطأ التقدير مشابه لتلك التي تكلمنا عنها في الفقرة أعلاه. مع تغير واحد وهو أن $S_i^2=P_iQ_i$ لذلك تصبح المعادلة كما يأتي

$$n \geq \frac{\sum_{i=1}^{K} N_i^2 P_i Q_i / W_i}{N^2 D + \sum_{i=1}^{K} N_i P_i Q_i}$$

حيث إن $D = d^2/4$.

مثال (6): في المثال (4) إذا علمنا من تجارب سابقة أن $P_1 = P_3 = 0.25$ وكذلك $P_2 = 0.5$ وإذا حددنا مقدار الخطأ في التقدير $d = 0.05$ أوجد حجم العينة

الحل:

$$n \geq \frac{\sum_{i=1}^{K} N_i^2 P_i Q_i / W_i}{N^2 D + \sum_{i=1}^{K} N_i P_i Q_i} = 319$$

9.6 تحديد عدد الطبقات

لتحديد عدد الطبقات هناك نقطتان يجب ملاحظتهما:

1. كيف سينقص $var(\bar{y}_{st})$ عندما تزداد عدد الطبقات.
2. كيف ستتأثر تكاليف المسح عندما يتغير عدد الطبقات.

النقطة الأولى نوقشت من قبل عدة باحثين منهم (1961) Cochran و Sethi (1963) و(1966) .Hess et. al وتوصلوا إلى عدة نماذج لتحديد عدد الطبقات. أما النقطة الثانية: لقد درس هذا الموضوع من قبل (1963) Sethi واقترح أن زيادة عدد الطبقات أكثر من 6 طبقات قلما يكون مفضلاً.

يمكن إثبات أن تقسيم المجتمع إلى طبقات يؤدي إلى زيادة الكفاءة إذا ارتفع عدد الطبقات. لكن الحد الأعلى لعدد الطبقات يكون بمقدار حجم العينة، بحيث نصل إلى النقطة التي لا نستطيع فيها أن نسحب سوى وحدة واحدة من كل طبقة. قد لا تكون هناك فائدة من زيادة عدد الطبقات في المسوحات الكبيرة إلى الحد الأعلى لأنه سيؤدي إلى زيادة التكاليف.

10.6 تقسيم المجتمع إلى طبقات بعد سحب العينة

في بعض الأحيان تظهر مشكلة عند المعاينة ألا وهي رغبتنا في تقسيم المجتمع إلى طبقات حسب متغير مهم، ولكن لا نستطيع وضع وحدات المعاينة في الطبقات المناسبة إلا بعد سحب العينة، على سبيل المثال ربما نرغب في تقسيم المجتمع بناء على الجنس لمعرفة آرائهم في قضية معينة، ونقوم بالمسح باستخدام الهاتف لكن الشخص الذي سيرد على الهاتف لا يمكن معرفة جنسه إلا بعد الاتصال به.

لنفترض أن عينة عشوائية بسيطة سحبت من أفراد المجتمع لغرض معرفة آرائهم في قضية معينة. هنا لا يمكن توزيع العينة بين الطبقتين إلى n_1 رجال و n_2 نساء إلا بعد مقابلة الأفراد الذين سحبوا في العينة. والآن بدلاً من استخدام $\bar{y}$ الوسط الحسابي للعينة العشوائية البسيطة كتقدير إلى الوسط الحسابي للمجتمع $\bar{Y}$ يمكن أن نستخدم $\bar{y}_{st}$ الوسط الحسابي للعينة الطبقية شريطة أن N_i/N معروفة للرجال والنساء. ولكن n_1 , n_2 في هذه الحالة متغيرات لأنهم يتغيرون من مسح إلى آخر حتى ولو كانت n، ثابتة لذلك فإن هذه العينة ليست بالضبط هي عينة عشوائية طبقية حسب تعريف العينة العشوائية الطبقية. ولكن إذا كان N_i/N معروفاً وإذا كانت $n_i > 20$ وحدة من كل طبقة، فإن هذه الطريقة لتقسيم المجتمع إلى طبقات بعد سحب العينة تكون قريبة من دقة المعاينة العشوائية الطبقية مع استخدام التوزيع المتناسب لتوزيع العينة بين الطبقات.

تمارين

1. ترغب إحدى الشركات في تقدير عدد ساعات العمل التي تضيع بسبب الحوادث التي يتعرض لها العاملون في الشركة، وبما أن العمال والفنيين والإداريين يختلف معدل الحوادث لكل فئة منهم لذلك قرر الباحث سحب عينة عشوائية طبقية بحيث إن كل فئة من العاملين الثلاثة تمثل طبقة. لدينا بيانات من دراسات سابقة تعطينا النتائج الآتية

الطبقة	S_i^2	N_i
العمال	36	132
الفنيون	25	92
الإداريون	9	27

إذا قرر الباحث سحب عينة عشوائية بحجم n=35 حدد توزيع العينة بين الطبقات باستخدام التوزيع الأمثل والتوزيع المتناسب.

2. قام الباحث بسحب عينة عشوائية بسيطة من كل طبقة للسؤال السابق، وحصل على عدد الساعات الضائعة خلال الشهر الماضي كما هي موضحة في الجدول الآتي:

الطبقة	عدد الساعات الضائعة
العمال	8, 0, 24, 20, 0,16,12, 0, 5, 6, 7, 8, 2, 5, 4, 3, 2
الفنيون	4, 5, 0, 8, 7, 6,1 , 3, 2, 0, 4,1 , 7, 24
الإداريون	8,1 , 0, 4

أ. قدر معدل عدد الساعات الضائعة خلال الشهر. وأوجد 90% فترة ثقة للوسط الحسابي.

ب. استخدم التوزيع الأمثل والمتناسب لتقدير عدد الساعات الضائعة وقارن بين النتائج التي تحصل عليها هنا ونتائج الفرع أعلاه.

ت. استخدم العينة العشوائية البسيطة لتقدير عدد ساعات العمل الضائعة وتقدير التباين، هل تعتقد

أن استخدام العينة الطبقية أفضل من العينة العشوائية البسيطة؟ وضح إجابتك.

3. لنفترض أن البيانات في السؤال الثاني أعلاه تمثل عينة عشوائية أولية، استخدمها لتحديد حجم العينة المطلوب، إذا كان حد الخطأ لتقدير الوسط الحسابي الطبقي $d=1.5$ ودرجة ثقة 95% .

4. استخدم البيانات المعطاة في السؤال الثاني لتحديد حجم العينة المطلوبة إذا كانت $d=1.5$ ودرجة الثقة 90%.

5. إذا كانت نسبة العاملين الذين تتجاوز أعمارهم 50 سنة من السؤال الثاني هي كما يأتي: العمال 20% والفنيون 20% والإداريون 35% قدر نسبة العاملين في الشركة الذين تزيد أعمارهم عن 50 سنة، ثم أوجد فترة 95% ثقة للنسبة الحقيقية.

6. يحتوي مجتمع على 4 طبقات، الجدول الآتي يعطينا معلومات عن الطبقات في المجتمع

N_i	$\bar{y}_i$	S_i^2	s_i^2	n_i	الطبقة
120	7.4	1.3	1.5	25	1
100	7	2.1	2.7	20	2
75	11	1.1	0.9	15	3
40	9.1	1.9	2.1	10	4

أ. أحسب $\bar{y}_{st}$ و Y_{st} و $var(\bar{y}_{st})$ و $var(Y_{st})$ و $s^2(\bar{y}_{st})$ و $s^2(Y_{st})$.

ب. استخدم التوزيع المتناسب والتوزيع الأمثل لتوزيع العينة بين الطبقات، ثم أوجد تباين الوسط الحسابي الطبقي للحالتين، ثم

قارن النتائج التي حصلت عليها هنا مع النتائج التي حصلت عليها في أ.

ت. استخدم البيانات التي لديك لتحديد حجم العينة المطلوب لتقدير المجموع الكلي إذا حدد مقدار حد الخطأ $d=400$ ودرجة الثقة95%.

7. لنفترض أن تكاليف معاينة الوحدة الواحدة في الطبقات الخمسة الواردة في السؤال السادس هي كما يأتي: $c_1 = 3, c_2 = 2.5, c_3 = 2, c_4 = 1.5$ استخدم التوزيع الأمثل لتوزيع حجم العينة بين الطبقات، وكذلك أوجد تقدير الوسط الحسابي الطبقي وتباينه.

8. إذا كانت تكاليف معاينة الوحدة الواحدة كما وردت في السؤال السابع وتقدير النسبة P_i للطبقة المختلفة هي كما يأتي: $\hat{P}_1 = 0.2, \hat{P}_2 = 0.3, \hat{P}_3 = 0.15, \hat{P}_4 = 0.12$ قدر النسبة الحقيقية باستخدام النسبة الطبقية $\hat{p}_{st}$ ثم أوجد معامل التغير إلى $\hat{p}_{st}$.

9. قام الباحثون الاقتصاديون في أحد البحوث بتصنيف القرى الموجودة في منطقة جغرافية معينة على أربع طبقات، بناءً على ارتفاع القرية عن مستوى سطح البحر. سحبت عينة عشوائية بسيطة حجمها 10 قرى (من كل طبقة) بما فيها القرى غير المأهولة بالسكان، وجرت عملية حصر عدد المساكن الموجودة في كل قرية، وحصلنا على النتائج المبينة في الجدول الآتي:

عدد المساكن في كل قرية ظهرت في العينة

الطبقة	عدد القرى	1	2	3	4	5	6	7	8	9	10
1	1400	43	84	98	0	10	44	0	124	13	0
2	4700	50	147	62	87	84	158	170	104	56	160
3	2500	228	262	110	232	139	178	334	0	63	220
4	14900	17	34	25	34	36	0	25	7	15	31

أ. قدر مجموع المساكن الموجودة في المنطقة وكذلك قدر التباين للمجموع الكلي .

ب. قارن التوزيع الحالي لحجم العينة بين الطبقات والتوزيع الأمثل لحجم العينة بين الطبقات.

ت. قدر الزيادة في الدقة الناتجة باستخدام العينة العشوائية الطبقية بالمقارنة بالعينة العشوائية البسيطة .

10. ترغب إحدى المدارس الابتدائية بتقدير علامات الطلبة في القراءة لطلبة الصف السادس الابتدائي، قامت المدرسة بتقسيم الطلاب إلى ثلاث مجموعات، المجموعة الأولى تحتوي على الطلبة سريعي التعلم، والمجموعة الثالثة على الطلبة بطيئي التعلم. تحتوي المجموعة الأولى على 55 طالباً والمجموعة الثانية على80، والمجموعة الثالثة على 65 طالباً. قامت المدرسة بسحب عشوائية طبقية بحجم 30 طالباً موزعة على المجموعات الثلاث، الجدول الآتي يعطينا نتائج الطلاب الذين سُحبوا من المجموعات (الطبقات) الثلاث.

المجموعة الأولى	7559 77 80 90 85 78 91
المجموعة الثانية	68 60 49 42 61 59 72 64 56 54 67 81
المجموعة الثالثة	19 50 42 44 56 67 33 45 47 61

أ. قدر معدل علامات الطلاب للصف السادس ومن ثم أوجد 95% فترة ثقة لمعدل علامات الطلاب الحقيقي.

ب. قدر الفرق بعلامات الطلاب بين المجموعة الأولى والثالثة ومن ثم أوجد 95% فترة ثقة للفرق الحقيقي بين المجموعتين.

11. لنفترض أن المدرسة ترغب في تقدير علامات الطلاب في نهاية العام الدراسي وتريد أن تسحب عينة جديدة وبحجم 30 طالباً ، وتوزعها بين

المجموعات الثلاث باستخدام الطريقة المثلى لتوزيع حجم العينة بين الطبقات. لنفترض أن تكلفة جمع البيانات من المجموعات (الطبقات) الثلاث متساوية، استخدم المعلومات في السؤال العاشر لتقدير التباينات.

12. استخدم البيانات في السؤال العاشر لتحديد حجم العينة لتقدير الوسط الحسابي للمجتمع ليكون بين ± 4، استخدم التوزيع المتناسب.

13. كرر العملية للسؤال الثاني عشر باستخدام التوزيع الأمثل.

14. ترغب إدارة المدرسة في تقدير نسبة الطلاب الذين حصلوا على علامات أقل من 60 أستخدم البيانات المعطاة في السؤال العاشر لتقدير نسبة الطلاب الحاصلين على علامة أقل من 60، ثم أوجد فترة 95% ثقة للنسبة الحقيقية للطلاب الحاصلين على أقل من 60.

15. يحتوي أحد المجتمعات على 8000 عنصر جرى تقسيمه إلى طبقتين تحتويان على التوالي 5000 و3000 عنصر. جرى سحب عينة عشوائية بسيطة من كلتا الطبقتين بواقع 200 و100 عنصر. الجدول الآتي يلخص النتائج التي حصلنا عليها.

$\hat{p}_i$	s_i^2	$\bar{y}_i$	n_i	N_i	الطبقة
0.25	50	400	200	5000	1
0.36	35	450	100	3000	2

أ. احسب $\bar{y}_{st}$ و $\hat{p}_{st}$ و $s(\bar{y}_{st})$ و $s(\hat{p}_{st})$.

ب. أوجد 95% فترة ثقة للوسط الحسابي $\bar{Y}$ والنسبة P.

16. لنفترض أن البيانات في السؤال الخامس عشر تمثل عينة عشوائية أولية، ونرغب أن نحدد حجم العينة وتوزيعها بين الطبقتين باستخدام التوزيع المتناسب بحيث يكون الوسط الحسابي والنسبة للمجتمع بحدود ∓ 50 و ± 0.7 على التوالي.

17. كرر السؤال السادس عشر ولكن باستخدام التوزيع الأمثل لتوزيع لحجم العينة بين الطبقات.

18. لنفترض أن تكاليف مشاهدة العنصر الواحد للطبقتين هي $C_1 = 9$ و$C_2 = 6$ استخدم التوزيع الأمثل والمعلومات المعطاة في السؤال الخامس عشر لتحديد حجم العينة وتوزيعها بين الطبقتين.

19. يبلغ عدد سكان إحدى المدن 500.000 نسمه حوالي 60% من البيض و25% من السود والباقي يمثلون 15% . لنفترض أن التباين في الطبقات الثلاث على النحو الآتي: $S_1^2 = 5000$ و $S_2^2 = 500$ و $S_3^2 = 200$. نرغب في سحب عينة بحجم n=200. استخدم التوزيع الأمثل لتوزيع العينة بين الطبقات الثلاث.

20. قدر نسبة الكراسي المكسورة في قاعات إحدى كليات جامعتك، كذلك قدر معدل سعة القاعة في الكلية.

21. صمم استبانة لاستطلاع رأي طلبة جامعتك في قضية معينة على اعتبار أن الكليات في الجامعة تمثل طبقات مختلفة، ثم نفذ المشروع للحصول على البيانات المطلوبة.

22. اقترح مشروعاً له علاقة بدراستك أو مجال عملك تستخدم فيه العينة العشوائية الطبقية لتقدير الوسط الحسابي للمجتمع والمجموع الكلي وقم بتنفيذه.

المراجع العربية

1. سليمان محمد طشطوش. (2001) أساسيات المعاينة الإحصائية، دار الشروق للنشر والتوزيع ، عمان.
2. حسين علوان (1994). طرق المعاينة، دار الفرقان، عمان.
3. هلال عبود البياتي وصبري رديف العاني (1981). العينات، جامعة بغداد، بغداد.
4. وليد عبدالحميد نوري وعبدالمجيد حمزة الناصر (1981). العينات، وزارة التعليم العالي والبحث العلمي، بغداد.
5. وليم كوكوران – ترجمة أنيس كنجو (1995). تقنية المعاينة الإحصائية، جامعة الملك سعود، الرياض.

References

1. Aoyama, H. (1954). A Study of the Stratified Random Sampling, *Ann. Inst. Statist. Math.* 6, 1–36.
2. Armitage, P. (1947). A Comparison of Stratified with Unrestricted Random Sampling from a finite Population. *Biometrika*, 34, 273–280/
3. Barnett, V. (1991). Sampling Survey Principles and Methods, Edward Arnold, London.
4. Bryant, E. C., Hartley, H. O. and Jessen, R. J. (1960). Design and Estimation in Two-way Stratification. *J. Amer. Statist. Assoc.* 55, 175–182.
5. Benedetto, J. J. and Ferreira, P. (2001). Modern Sampling Theory, Birkhauser Boston.
6. Chatterjee, S. (1967). A Note on Optimum Stratification. *Skand. Akt.* 50, 40–44.
7. Chatterjee, S. (1968). Multivariate Stratified Survey. *J. Amer. Stat. Asso.* 63, 530–534.
8. Chaudhuri, A. and Stenger, H. (1992). Survey Sampling: Theory and Methods, Marcel Dekker, New York.
9. Cochran, W. G. (1961). Comparison of Methods for Determined Stratum Boundaries. *Bull. Inst. Stat. Inst.*, 38, 345–358.
10. Cochran, W. G. (1977). Sampling Technique, 3rd Ed. Wiley, New York.
11. Cornell, F. G. (1947). A Stratified Random Sample of a Small Finite Population. *J. Amer. Stat. Asso.* 42, 523–532..
12. Dalenius, T. and Gurney, M. (1951). The Problem of Optimum Stratification II. *Skand. Aktuartidsk* 7–34, 133–148.
13. Dalenius, T. and Hodges, J. L. (1957). The Choice of Stratification Points. *Skand. Aktuar-tidskr.* 3–4, 198–203.
14. Dalenius, T. and Hodges, J. L. (1959). Minimum Variance Stratification. *J. Amer. Statist. Assoc.* 54, 88–101, correction JASA, 58 (1963), p. 1161.
15. Deming, W. E. (1960). Sampling Design in Business Research. Wiley, New York
16. Ekman, G. (1959). An Approximation useful in Univariate Stratification. *Ann. Math.Statist.* 30,219–229.

17. Evans, W. D. (1951). On Stratification and Optimum Allocations. *J. Amer. Statist. Assoc.* 46, 95□104.
18. Foreman, E. K. (1991). Survey Sampling Principles, Marcel Dekker, New York.
19. Fuller, W. A. (1966). Estimation Employing Post Strata. *J. Amer. Statist. Assoc.* 61, 1172□1183.
20. Fuller, W. A. (1970). Sampling with Random Stratum Boundaries. *J. Roy. Stat. Soc.* B32, 209□206.
21. Govindarajulu, Z. (1999). Elements of Sampling Theory and Methods, Prentice Hall.
22. Goodman, R. and Kish, L. (1950). Controlled Selection□a Technique in Probability Sampling. *J. Amer. Statist. Assoc.*, 45, 350□372.
23. Hajek, J (1981). Sampling from Finite Population, Marcel Dekker, New York.
24. Hansen, M. H., and Hurwitz, W. N. (1943). On the Theory of Sampling from Finite Populations. *Ann. Math. Statist.* 14, 333□362.
25. Hedayat, A. S. and Sinha, B. K. (1991). Design and Inference in Finite Population Sampling, Wiley, New York.
26. Hess, I., Sethi, V. K., and Balakrishnan, T. R. (1966). Stratification□ A Practical Investigation. *J. Amer. Statist. Assoc.* 61, 74□90.
27. Holt, D. and Smith, T. M. F. (1979). Post□stratification, *J. Roy. Statist. Soc.* A142, 33□46.
28. Kish, L (1995). Survey Sampling, Wiley, New York.
29. Levy, P. S. and Lemeshow S. (1991). Sampling of Population□Methods and Applications, Wiley, New York.
30. Lohr, S. L. (1999). Sampling□Design and Analysis, Duxbury, New York.
31. Mahalanobis, P. C. (1946). Recent Experiments in Statistical Sampling in the Indian Statistical Institute. *J. Roy. Stat. Soc.* A109, 325□378, reprinted in Sankhyii
32. Murthy, M. N. (1962). Variance and Confidence Interval Estimation. Sankhyii 24(B), 1□12.
33. Neyman, J. (1934). On the Two Different Aspects of the Representative Method□ The Method of Stratified Sampling and the Method of Purposive Selection, *J. Roy. Stat. Soc.* 97, 558□606.
34. Raj, D. (1968). Sampling Theory, McGraw Hill, New York.
35. Rao, J. N. K. (1985). Conditional Inference in Survey Sampling. Survey Methodology, D. I. *Inst. Statist. Math.* No.1, 15□31.
36. Sampath, S. (2000). Sampling Theory and Methods, CRC Press.
37. Satterthwaite, F. E. (1946). An Approximate Distribution of Estimates of Variance Components, *Biometrics* 2, 110□114.
38. Scheaffer, R.L., Mendenhall, W. and Ott, L. (1996). Elementary Sampling Survey, 5th ed., Duxbury, New York.
39. Serfling, R. J. (1968). Approximately Optimal Stratification, *J. Amer. Statist. Assoc.* 63, 1298□1309.
40. Sethi, V. K. (1963). A Note on Optimum Stratification of Populations for Estimating the Population Means, *Aust. J. Statist.* 5, 20□33.
41. Sirken, M. G. (1972). Stratified Sample Surveys with Multiplicity, *J. Amer. Statist. Assoc.* 67, 224□227.
42. Smith, T.M.F. (1976). The Foundation of Survey Sampling□A review, *Journal of Royal Statistics Society* A139, p183□204.

43. Stephan, F. F. (1945). The Expected Value and Variance of the Reciprocal and other Negative Powers of a Positive Bernoulli Variate. *Ann. Math. Statist.* 16, 50-61.
44. Sturat A. (1984). The Ideas of Sampling, Revised Edition, Griffin, London.
45. Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., and Ashok, C. (1984). Sampling Theory of Surveys with Applications, 3rd ed., Ames (Iowa) Iowa State University Press.
46. Taga, Y. (1967). On Optimum Stratification for the Objective Variable based on Concomitant Variables using Prior Information. *Ann. Inst. Statist. Math.* (Tokyo) 4, 95-102.
47. Thompson, S. K (2002). Sampling, 2end Wiley, New York.
48. Tryfos, P. (1996). Sampling Methods for Applied Research, Wiley, New York.
49. Yates, F. (1981). Sampling Methods for Censuses and Surveys 4th Ed., Grittin, London

الفصل السابع

التقدير بطريقة النسبة والانحدار والفرق
Ratio, Regression and Difference Estimation

1.7 مقدمة

قد تظهر في كثير من المسوحات معلومات مساعدة أو إضافية لمتغير له علاقة مع المتغير الذي نحن بصدد دراسته. هذه المعلومات يمكن الاستفادة منها لتحسين تصميم المعاينة. ومن ثم الوصول إلى تقديرات أكثر دقة لمعلمات المجتمع للمتغير الذي ندرسه. فمثلاً إذا كنا نقوم بمسح لتقدير مصروفات العائلة الشهرية لمدينة معينة.. فإن عدد أفراد العائلة له علاقة مباشرة وقوية مع مصروفات العائلة الشهرية. غالباً نستطيع الحصول على معلومات عن عدد أفراد العائلة في المجتمع من دوائر الأحوال المدنية أو من مسوحات شاملة جرى فيها جمع معلومات كاملة عن عدد أفراد العائلة وكذلك معلومات أخرى عن العائلة.

إن المعلومات الإضافية أو المساعدة التي يمكن الاستفادة منها تكون متوافرة على شكل مجاميع كلية أو معدلات، وليس عن كل وحدة من وحدات المجتمع؛ لذلك يمكن استخدام هذه المعلومات عند القيام بالمسح بالعينة، شريطة أن تكون هذه المعلومات الخاصة بالمتغير المساعد يمكن الحصول عليها بسهولة عند جمع معلومات حول المتغير الرئيس باستخدام العينة. ففي مثالنا الخاص بتقدير مصروفات العائلة الشهرية لا توجد صعوبة

عنـد سـحب عينـة مـن العـائلات ومقـابلتهم أن نسـألهم بالإضـافة إلى مقـدار مصروفاتهم الشهرية عن عدد أفراد العائلة.

27 رموز ومصطلحات

إن هـذه الرمـوز والمصـطلحات الـتي سـوف نعرفهـا هنـا خاصـة بالعينـة العشوائية البسيطة. أما الرموز والمصطلحات الخاصة بالعينة العشوائية الطبقية فلا توجد صعوبة في الحصول عليها. وسوف نقوم بتعريف ما نحتاجه منها لاحقاً.

y_i : قيمة المتغير أو الصفة تحت الدراسة للوحدة i في المجتمع.

x_i : قيمة المتغير أو الصفة المساعدة في نفس الوحدة i في المجتمع.

Y : المجموع الكلي للمتغير y في المجتمع.

X : المجموع الكلي للمتغير x في المجتمع.

$R = \frac{\bar{Y}}{\bar{X}} = \frac{Y}{X}$: نسبة المجموع والوسط للمتغير y و x.

ρ : معامل الارتباط بين y و x في المجتمع.

لنفترض أننا نرغب في تقدير المجموع Y أو الوسط $\bar{Y}$ للمجتمع. وسحبنا عينة عشوائية بسيطة بحجم n من وحدات المجتمع. وحصلنا على n زوج من المشاهدات حول المتغيرين y و x. ولنفترض أن $\bar{y}$ و $\bar{x}$ الوسط الحسابي للعينة للمتغيرين y و x على التوالي، فعليه يكون R تقدير للنسبة R

$$R = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} = \frac{\bar{y}}{\bar{x}}$$

و $\bar{y}_R$ تقدير الوسط الحسابي للمجتمع بطريقة النسبة

$$\bar{y}_R = \frac{\bar{y}}{\bar{x}} \bar{X} = R\bar{X}$$

و Y_R تقدير المجموع الكلي للمجتمع بطريقة النسبة

$$Y_R = \frac{\bar{y}}{\bar{x}} X = R X$$

3.7 تقدير النسبة R في العينة العشوائية البسيطة

قد يحدث في كثير من الدراسات أن نجمع معلومات عن أكثر من متغير في الوقت نفسه. باستمرار توجد رغبة في دراسة العلاقة بين هذه المتغيرات، بالإضافة إلى دراسة هذه المتغيرات على حدة. إن النسبة بين متغيرين R من العلاقات التي نرغب في دراستها باستمرار. على سبيل المثال نرغب أن نعرف نسبة دخل الفرد أو العائلة المصروف على الطعام، أو نسبة الإنتاج التالف لأحد المصانع.

لنفترض أن المتغير الذي نرغب في دراسته y والمتغير الإضافي أو المساعد x فإن النسبة لهذين المتغيرين في المجتمع تسمى نسبة المجتمع للمتغيرين وتعرف

$$R = \frac{\bar{Y}}{\bar{X}} = \frac{Y}{X}$$

لتقدير هذه النسبة سحبنا عينة عشوائية بسيطة حجمها n وحدة من المجتمع، وجمعنا معلومات حول المتغيرين y وx، ولنفترض أن $\bar{y}$ و $\bar{x}$ هما الوسط الحسابي للعينة للمتغيرين y وx على التوالي. نقدر R باستخدام R والتي سبق تعريفها أعلاه. إن R هو تقدير متحيز إلى R أي إن $E(R) \neq R$. لمزيد من المعلومات يراجع حسين علوان (1993) و(1977) Cochran أو ترجمة أنيس كنجو (1995). ننتقل الآن لإيجاد تباين تقدير النسبة، إذا كان حجم العينة كبيراً فإن تباين R سيكون تقريباً

$$Var(R) = \frac{1 - f}{n\bar{X}^2}[S_y^2 + R^2 S_x^2 - 2RS_{xy}]$$

حيث إن

$$S_x^2 = \frac{\sum_{i=1}^{N}(X_i - \overline{X})^2}{N-1}$$

و

$$S_y^2 = \frac{\sum_{i=1}^{N}(Y_i - \overline{Y})^2}{N-1}$$

و

$$f = \frac{n}{N}$$

و

$$S_{xy} = \frac{\sum_{i=1}^{N}(Y_i - \overline{Y})(X_i - \overline{X})}{N-1}$$

لا بد من تقدير $var(R)$ لأن R و S_x^2 و S_y^2 و S_{xy}^2 غير معلومات. لذلك فإن تقدير تباين $var(R)$ يمكن إيجاده باستخدام:

$$s^2(R) = \frac{1-f}{n\bar{x}^2}[s_y^2 + R^2 s_x^2 - 2R\, s_{xy}]$$

حيث إن

$$s_x^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

و

$$s_y^2 = \frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-1}$$

و

$$s_{xy} = \frac{\sum_{i=1}^{n}(y_i - \bar{y})(x_i - \bar{x})}{n-1}$$

إن $s^2(R)$ تقدير متحيز إلى $var(R)$ ولكن إذا كان حجم العينة كبيراً فإن قيمة التحيز تكون قليلة. بقي أن نكتب $s^2(R)$ بشكل آخر يسهل استخدامه في العمليات الحسابية:

$$s^2(R)=\frac{1-f}{n\bar{x}^2}\;\frac{\sum_{i=1}^{n}y_i^2+R^2\sum_{i=1}^{n}x_i^2-2R\sum_{i=1}^{n}y_i\,x_i}{n-1}$$

مثال (1): الجدول الآتي يمثل عدد الأسماك التي تم اصطيادها في بحيرة. وعدد الصيادين الذين استخدموا البحيرة مدة 20 يوماً. إذا كان N=200 يمثل عدد الأيام المسموح فيها الصيد خلال السنة.

اليوم	1	2	3	4	5	6	7	8	9	10
عدد الصيادين x_i	4	5	2	5	1	1	7	6	4	1
عدد الأسماك y_i	8	10	8	15	2	3	18	15	20	4
اليوم	11	12	13	41	15	16	17	81	19	20
عدد الصيادين x_i	2	4	1	2	3	4	4	2	0	2
عدد الأسماك y_i	6	14	3	7	8	10	12	4	0	7

استخدم هذه البيانات لتقدير نسبة عدد الأسماك إلى عدد الصيادين ثم أوجد تقدير معامل التغير إلى تقدير النسبة.

الحل:

$$\sum_{i=1}^{20}x_i\,y_i=697 \text{ و } \sum_{i=1}^{20}x_i^2=248 \text{ و } \sum_{i=1}^{20}y_i^2=2094 \text{ و } \sum_{i=1}^{20}x_i=60 \text{ و } \sum_{i=1}^{20}y_i=174$$

فعليه يكون تقدير النسبة

$$R = \frac{\bar{y}}{\bar{x}} = \frac{8.7}{3} = 2.9$$

و تقدير تباين النسبة

$$s^2(R) = \frac{N-n}{Nn\bar{x}^2} \; \frac{\sum_{i=1}^{n} y_i^2 + R^2 \sum_{i=1}^{n} x_i^2 - 2R \sum_{i=1}^{n} y_i x_i}{n-1}$$

$$= \frac{(200-20)}{20(3)^2} \; \frac{2094 + (2.9)^2(248) - 2(2.9)(697)}{(200)(19)} = 0.0361$$

إن تقدير الخطأ المعياري لتقدير النسبة

$$s(R) = \sqrt{s^2(R)} = \sqrt{0.0361} = 0.19$$

كذلك فإن معامل التغير لتقدير النسبة

$$Cv(R) = \frac{s(R)}{R} = \frac{0.19}{2.9} = 0.0655$$

إذا كان X المجموع الكلي للمتغير x معلوماً فإن تقدير النسبة يصبح

$$R = \frac{Y}{X}$$

وهذا تقدير غير متحيز إلى R وتباينه

$$Var(R) = \frac{1-f}{\bar{X}^2} \; \frac{S_y^2}{n}$$

لمزيد من المعلومات يراجع حسين علوان (1993) و(1977) Cochran أو ترجمة أنيس كنجو (1995).

4.7 تقدير الوسط الحسابي والمجموع باستخدام النسبة

إن الوسط الحسابي للمجتمع $\bar{Y}$ ومجموع المجتمع الكلي Y يمكن تقديرهما باستخدام تقدير النسبة على النحو الآتي:

تقدير الوسط الحسابي بطريقة النسبة

$$\bar{y}_R = R\bar{X}$$

تقدير المجموع بطريقة النسبة

$$Y_R = RX$$

نفترض معرفة الوسط الحسابي $\bar{X}$ والمجموع الكلي X للمجتمع للمتغير x واضح أن $\bar{y}_R$ و Y_R تقديران متحيزان إلى $\bar{Y}$ و Y على التوالي مع تباين تقريبي للتقديرين $\bar{y}_R$ و Y_R على التوالي

$$Var(\bar{y}_R) = \bar{X}^2 Var(R) = \frac{1-f}{n}[S_y^2 + R^2 S_x^2 - 2RS_{xy}]$$

$$Var(Y_R) = N^2 Var(\bar{y}_R) = \frac{1-f}{n} N^2 [S_y^2 + R^2 S_x^2 - 2RS_{xy}]$$

وتقدير التباين للتقديرين $\bar{y}_R$ و Y_R هما على التوالي

$$s^2(\bar{y}_R) = \frac{(1-f)}{n} \frac{\sum_{i=1}^{n} y_i^2 + R^2 \sum_{i=1}^{n} x_i^2 - 2R \sum_{i=1}^{n} y_i x_i}{n-1}$$

$$s^2(Y_R) = N^2 s^2(\bar{y}_R)$$

باستطاعتنا المقارنة بين تباين الوسط الحسابي للعينة العشوائية البسيطة $var(\bar{y})$ وتباين الوسط الحسابي للعينة العشوائية البسيطة باستخدام النسبة $var(\bar{y}_R)$ لنجد أن

$$var(\bar{y}_R) \le var(\bar{y})$$

عندما تكون

$$2RS_{xy} \geq R^2 S_x^2$$

أو

$$\rho \geq \frac{1}{2}\frac{CV(x)}{CV(y)}$$

تشير هذه النتائج إلى أن تقدير الوسط الحسابي بطريقة النسبة سيكون أكثر دقة من استخدام المتغير y بمفرده، إذا كان معامل الارتباط ρ أكبر من معامل تغير X على معامل تغير y .

مثال (2): في المثال (1) لنفترض أن $X=650$ عدد الصيادين في السنة. أوجد تقدير الوسط الحسابي بطريقة النسبة $\bar{y}_R$ وبطريقة y بمفرده وقارن بينهما.

الحل:

$$s^2(\bar{y}_R) = \frac{(N-n)}{Nn}\;\frac{\sum_{i=1}^{n} y_i^2 + R^2\sum_{i=1}^{n} x_i^2 - 2R\sum_{i=1}^{n} y_i x_i}{n-1}$$

$$= \frac{200-20}{20(200)}\;\frac{2094 + (2.9)^2(248) - 2(2.9)(697)}{19} = 0.3247$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{20}(174) = 8.7$$

$$s^2(\bar{y}) = \frac{N-n}{Nn}\,s_y^2 = \frac{200-20}{200}\;\frac{30.54}{20} = 1.3743$$

مما لا شك فيه لقد طرأ تحسن كبير على دقة التقدير للوسط الحسابي للمجتمع باستخدام النسبة.

5.7 اختيار حجم العينة

إن مقدار المعلومات المحتواة في العينة يعتمد على عاملين، الأول التشتت في البيانات، وهو الذي يمكن السيطرة عليه في بعض الأحيان وذلك من خلال تصميم المسح بالعينة. وأما العامل الثاني فهو حجم العينة n وبعد اختيار وتصميم العينة لا بد من تحديد حجم العينة. وسوف نناقش اختيار حجم العينة المطلوبة لتقدير معلمات المجتمع النسبة R والوسط الحسابي والمجموع الكلي للمجتمع.

1.5.7 اختيار حجم العينة لتقدير النسبة R

إن الطريقة التي سوف تتبع هنا لاختيار حجم العينة مماثله لما سبق في فصول سابقة. إن عدد الوحدات المطلوبة لتقدير نسبة المجتمع R ضمن حد معين لمقدار الخطأ في التقدير وليكن d ومع 95% ثقة إي

$$p(\left| R - R \right| < d) = 0.95$$

لذلك يجب حل المعادلة الآتية لإيجاد قيمة n

$$2\sqrt{var(R)} = d$$

سوف نستخدم $s^2(R)$ بدلاً من $var(R)$ لأن R و S_x^2 و S_y^2 و S_{xy}^2 غير معلومات.
إن حجم العينة المطلوب سيكون

$$n > \frac{Ns^2}{ND + s^2}$$

حيث إن $D = \frac{d^2 \overline{X}^2}{4}$. ولكننا في الواقع لا نعرف قيمة σ^2. إذاً لا بد من تقديرها من معلومات سابقه لمسوحات مشابه أو بسحب عينة أولية حجمها n' ثم نستخدمها لتقدير σ^2 بما يلي

$$\sigma^2 = \frac{1}{n'-1} \sum_{i=1}^{n'} (y_i - R\, x_i)^2$$

ونعوض $\hat{\sigma}^2$ بدلاً عن σ^2 في المعادلة أعلاه، وكذلك إذا كانت قيمة $\overline{X}$ غير معلومة نقدرها باستخدام $\overline{x}$ الوسط الحسابي للعينة الأولية.

2.5.7 اختيار حجم العينة لتقدير الوسط الحسابي للمجتمع باستخدام النسبة

يمكن أن نحدد حجم العينة الذي نحتاجه لتقدير الوسط الحسابي للمجتمع باستخدام النسبة ضمن حد معين لمقدار الخطأ في التقدير وليكن d وبثقة 95% أي

$$p(\left|\overline{Y} - \overline{y}_R\right| < d) = 0.95$$

وذلك بحل إحدى المعادلتين

$$2\sqrt{Var(\overline{y}_R)} = d$$

أو

$$2\overline{X}\sqrt{Var(R)} = d$$

إذاً حجم العينة المطلوب لتقدير الوسط الحسابي للمجتمع مع خطأ في التقدير مقداره d وبثقة 95% فإن حجم العينة المطلوب يكون

$$n > \frac{Ns^2}{ND + s^2}$$

حيث إن $D = \frac{d^2}{4}$. نلاحظ هنا أننا لا نحتاج إلى تقدير $\overline{X}$ ولكن لا بد من تقدير σ^2 بالطريقة نفسها التي استخدمناها في تقدير حجم العينة للنسبة.

3.5.7 اختيار حجم العينة لتقدير المجموع الكلي للمجتمع باستخدام النسبة

لا نحتاج لإعادة الكلام نفسه الذي استخدمناه عند الكلام عن اختيار حجم العينة لتقدير R و $\overline{Y}$ أعلاه. ونقول إن حجم العينة المطلوب يكون

$$n > \frac{Ns^2}{ND + s^2}$$

حيث إن

$$. D=\frac{d^2}{4N^2}$$

مثال (3): لنفترض في مثالنا (1) أن البيانات المعطاة في الجدول تمثل عينه أولية أي إن $n'=20$ وكذلك لنفترض أن $N=200$ و$X=650$. أوجد حجم العينة المطلوب لتقدير R و $\overline{Y}$ و Y إذا كانت قيمة d تساوي 0.3 و0.8 و100 على التوالي.

الحل: نقدر قيمة σ^2 باستخدام العينة الأولية

$$\sigma^2=\frac{1}{n'-1}\sum_{i=1}^{n'}(y_i-R\,x_i)^2=\frac{137.08}{19}=7.2147$$

إذاً حجم العينة المطلوب لتقدير R

$$n=\frac{N\sigma^2}{ND+\sigma^2}=\frac{N\sigma^2}{N(d^2\overline{X}/4)+\sigma^2}=\frac{200(7.2147)}{200[(0.3)^2(3.25)^2/4]+7.2147}=27$$

وكذلك فإن حجم العينة المطلوب لتقدير $\overline{Y}$

$$n=\frac{N\sigma^2}{ND+\sigma^2}=\frac{N\sigma^2}{N(d^2/4)+\sigma^2}=\frac{200(7.2147)}{200[(0.8)^2/4]+7.2147}=37$$

وأخيراً فإن حجم العينة المطلوب لتقدير Y يكون

$$n=\frac{N\sigma^2}{ND+\sigma^2}=\frac{N\sigma^2}{N(d^2/N^2 4)+\sigma^2}=\frac{200(7.2147)}{200[(100)^2/4(200)^2]+7.2147}=74$$

6.7 تقديرات النسبة في العينة العشوائية الطبقية

لقد لاحظنا من خلال دراستنا إلى الآن أن تقدير الوسط الحسابي والمجموع الكلي للمجتمع يمكن تحسينهما باستخدام العينة العشوائية

الطبقية. وأقصد بالتحسين تقليل التباين. وكذلك لاحظنا أن استخدام النسبة لتقدير الوسط الحسابي والمجموع للمجتمع يمكن أن يؤدي إلى تحسين التقدير لهذه المعلمات. والسؤال الذي يطرح نفسه ماذا لو استخدمنا النسبة عندما تكون العينة سحبت من مجتمع مقسم إلى طبقات؟ وهذا ما سنحاول دراسته هنا.

1.6.7 تقدير المجموع الكلي للمجتمع باستخدام النسبة

هناك طريقتان لتقدير المجموع الكلي باستخدام النسبة في العينة العشوائية الطبقية.

- **المجموع الطبقي النسبي المنفصل ويعرف**

$$Y_{RS} = \sum_{i=1}^{K} \frac{\bar{y}_i}{\bar{x}_i} X_i = \sum_{i=1}^{K} R_i X_i$$

حيث إن R_i و X_i يمثلان تقدير النسبة والمجموع الكلي للطبقة i و $i=1,2,\ldots,K$ للمتغير X. ولا بد من معرفة X_i لجميع الطبقات. كذلك فإن مقدار التحيز للتقدير سيكون عبارة عن مجموع التحيزات لكل الطبقات وسيحمل التحيز على الأغلب الإشارة نفسها لجميع الطبقات. أما تباين Y_{RS} وتقدير تباينه فسيكونان على التوالي

$$Var(Y_{RS}) = \sum_{i=1}^{K} \frac{N_i^2}{n_i}(1-f_i)[S_{yi}^2 + R_i^2 S_{xi}^2 - 2R_i S_{xyi}]$$

$$s^2(Y_{RS}) = \sum_{i=1}^{K} \frac{N_i^2}{n_i}(1-f_i)[s_{yi}^2 + R_i^2 s_{xi}^2 - 2R_i s_{xyi}]$$

2 المجموع الطبقي النسبي المشترك ويعرف

$$Y_{RC} = \frac{Y_{st}}{X_{st}} X = \frac{\sum_{i=1}^{K} N_i \bar{y}_i}{\sum_{i=1}^{K} N_i \bar{x}_i} X$$

حيث إن Y_{st} و X_{st} تقدير المجموع بالعينة العشوائية الطبقية للمتغيرين y و x و X يمثل المجموع الكلي للمتغير x.

يلاحظ أن Y_{RC} فقط يحتاج المجموع الكلي X وليس المجموع الكلي لكل طبقه. كذلك لأن النسبة لا نجدها حتى الخطوة الأخيرة سيكون مقدار التحيز أقل مما هو عليه في Y_{RS}. أما التباين وتقدير التباين للتقدير فسيكونا على التوالي:

$$Var(Y_{RC}) = \sum_{i=1}^{K} \frac{N_i^2}{n_i}(1-f_i)[S_{yi}^2 + R^2 S_{xi}^2 - 2R\, S_{xyi}]$$

$$s^2(Y_{RC}) = \sum_{i=1}^{K} \frac{N_i^2}{n_i}(1-f_i)[s_{yi}^2 + R^2 s_{xi}^2 - 2R\, s_{xyi}]$$

وهذا يختلف عن تباين Y_{RS} ، هناك $R_i = \frac{Y_i}{X_i}$ نستخدم، أما هنا فنستخدم $R = \frac{Y}{X}$. وهذا هو الذي غير النتيجة لصالح التقدير النسبي المنفصل ليكون التقدير النسبي المشترك أقل دقة من التقدير النسبي المنفصل.

2.6.7 تقدير الوسط الحسابي للمجتمع باستخدام النسبة

كما هو الحال في تقدير المجموع يوجد لدينا تقديران لتقدير الوسط الحسابي في العينة العشوائية الطبقية وباستخدام النسبة.

1. **الوسط الحسابي الطبقي المنفصل بطريقة النسبة ويعرف**

$$\bar{y}_{RS} = \frac{1}{N} Y_{Rs} = \frac{1}{N}\sum_{I=1}^{K} \frac{\bar{y}_i}{\bar{x}_i} X_i = \sum_{I=1}^{K} W_i \frac{\bar{y}_i}{\bar{x}_i} \bar{X}_i = \sum_{i=1}^{K} W_i \bar{y}_{Ri}$$

أما التباين وتقدير التباين لهذا التقدير $\bar{y}_{RS}$ فسيكونا

$$var(\bar{y}_{RS}) = \frac{1}{N^2} Var(Y_{RS})$$

$$s^2(\bar{y}_{RS}) = \frac{1}{N^2} s^2(Y_{RS})$$

2. **الوسط الحسابي الطبقي المشترك بطريقة النسبة ويعرف**

$$\bar{y}_{RC} = \frac{1}{N} Y_{RC} = \frac{1}{N} \frac{\sum_{i=1}^{K} N_i \bar{y}_i}{\sum_{i=1}^{K} N_i \bar{x}_i} X = \frac{\sum_{i=1}^{K} W_i \bar{y}_i}{\sum_{i=1}^{K} W_i \bar{x}_i} \bar{X}$$

أما التباين وتقدير التباين للتقدير $\bar{y}_{RC}$ فيكونا

$$var(\bar{y}_{RC}) = \frac{1}{N^2} var(Y_{RC})$$

$$s^2(\bar{y}_{RC}) = \frac{1}{N^2} s^2(Y_{RC})$$

مثال (4): استخدم الجدول الآتي لتقدير $\bar{Y}$ وY باستخدام النسبة في العينة العشوائية الطبقية. أوجد تقدير الخطأ المعياري للتقديرات وقارنها مع النتائج التي نحصل عليها باستخدام العينة الطبقية بمفردها أي بدون استخدام النسبة.

s_{xyi}	s^2_{xi}	s^2_{yi}	X_i	$\bar{x}_i$	$\bar{y}_i$	n_i	W_i	N_i	الطبقة
24	4	225	21 000	32	50	40	0.2	700	1
8.4	2.25	64	31 500	18.5	30	50	0.5	1750	2
4.32	0.64	36	8400	7	20	40	0.3	1050	3

$$Y_{RS}=\sum_{i=1}^{K}\frac{\bar{y}_i}{\bar{x}_i}X_i=\sum_{i=1}^{K}R_i\,X_i=107893$$

$$Y_{RC}=\frac{Y_{st}}{X_{st}}X=\frac{\sum_{i=1}^{K}N_i\,\bar{y}_i}{\sum_{i=1}^{K}N_i\,\bar{x}_i}X=\frac{108500}{62125}(60900)=106361$$

إذا أهملنا معامل التصحيح أي إن $(1 \square f_i)=1$ فإن

$$s^2(Y_{RS})=\sum_{i=1}^{K}\frac{N_i}{n_i}(1 \square f_i)[s_{yi}^2+R_i^2 s_{xi}^2 \square 2R_i\,s_{xyi}]=5026730$$

$$s\,(Y_{RS})=\sqrt{s^2(Y_{RS})}=\sqrt{5026730}=2242$$

$$s^2(\,Y_{RC})=\sum_{i=1}^{K}\frac{N_i}{n_i}(1 \square f_i)[s_{yi}^2+R^2 s_{xi}^2 \square 2R\ \ s_{xyi}]=5048091$$

$$s(Y_{RC})=\sqrt{s^2(\,Y_{RC})}=\sqrt{5048091}=2247$$

$$s^2(\bar{y}_{RS})=\frac{1}{N^2}s^2(Y_{RS})=0.641 \square\ \bar{y}_{RS}=\sum_{i=1}^{3}W_i\,\bar{y}_{Ri}=30.83$$

$$\square\ s^2(\bar{y}_{RC})=\frac{1}{N^2}s^2(Y_{RC})=0.642\ \bar{y}_{RC}=\frac{\sum_{i=1}^{K}W_i\,\bar{y}_i}{\sum_{i=1}^{K}W_i\,\bar{x}_i}\bar{X}=30.39$$

وإذا استخدمنا المتغير y بمفرده في العينة العشوائية الطبقية

فإن تقدير المجموع

$$Y_{st}=108500$$

وتقدير التباين

$$s^2(Y_{st})=7361025$$

نلاحظ أن هناك زيادة في الدقة عندما استخدمنا النسبة في التقدير. وكذلك نلاحظ أن التقدير النسبي المنفصل أكثر دقة من التقدير النسبي المشترك.

7.7 التقديرات باستخدام الفرق في العينة العشوائية البسيطة

توجد طريقة أخرى للاستفادة أو لاستخدام المعلومات الإضافية أو المساعدة لتقدير الوسط الحسابي والمجموع الكلي للمجتمع، تسمى الفروق، وباستخدام هذه الطريقة يمكن أن نعرف تقدير الوسط الحسابي

$$\bar{y}_d = \bar{y} + (\bar{X} - \bar{x})$$

إذا كان الوسط الحسابي للمجتمع والعينة للمتغير X هما $\bar{x}$ و $\bar{X}$ على التوالي و $\bar{y}$ هو الوسط الحسابي للعينة للمتغير y. يمكن ملاحظة أن $\bar{y}_d$ هو تقدير غير متحيز إلى $\bar{Y}$ على شرط أن يكون $\bar{y}$ و $\bar{x}$ تقديرين غير متحيزين للوسطين الحسابيين للمجتمع $\bar{Y}$ و $\bar{X}$ على التوالي.

إن هذا التقدير لا يخلو من صعوبات ويجب استخدامه بحذر. بصورة عامة معلومات المتغير X يجب أن تكون من النوع نفسه والوحدات نفسها كما هي للمتغير y. فمثلاً يمكن أن يكون x يمثل معلومات سريعة عن إنتاج محصول معين وجد عن طريق تقدير سريع لإنتاج هذا المحصول لجميع الحقول، بينما y يمثل معلومات دقيقة وجدت عن إنتاج المحصول نفسه وذلك عن طريق سحب عينة عشوائية من الحقول لتقدير إنتاج هذا المحصول.

إن $\bar{y}_d$ هو تقدير غير متحيز إلى $\bar{Y}$ كما أشرنا إلى ذلك. أما تباين وتقدير التباين إلى $\bar{y}_d$ فهما على التوالي

$$var(\bar{y}_d) = \frac{1-f}{n}(S_y^2 + S_x^2 - 2S_{xy})$$

$$s^2(\bar{y}_d) = \frac{1-f}{n}(s_y^2 + s_x^2 - 2s_{xy}) = \frac{1-f}{n}\frac{\sum_{i=1}^{n}[(y_i - x_i) - (\bar{y} - \bar{x})]^2}{n-1}$$

مثال (5): لأجل القيام بتضمين مزارع الخضراوات يقوم أشخاص لهم خبرة في موضوع تخمين قيمة هذه المزارع. يوجد 180 مزرعة قدرت قيمتها بمبلغ 13320 دولاراً. لنفترض أن x_i ترمز إلى القيمة التخمينية للمزرعة و y_i ترمز إلى القيمة الحقيقية للمزرعة. سحبنا عينه عشوائية حجمها $n = 10$ مزارع أعطتنا النتائج المدونة في الجدول أدناه. استخدم هذه البيانات لتقدير المجموع الكلي لقيمة هذه المزارع باستخدام طريقة الفرق، ثم أوجد معامل التغير.

المزرعة	1	2	3	4	5	6	7	8	9	10
y_i	9	14	7	29	45	109	40	238	60	170
x_i	10	12	8	26	47	112	36	240	59	167

الحل:

$$\bar{y} = 72.1 ,\ \bar{x} = 71.7 ,\ \bar{X} = 74$$

$$\bar{y}_d = \bar{y} + (\bar{X} - \bar{x}) = 72.1 + (74 - 71.7) = 74.4$$

$$s^2(\bar{y}_d) = \frac{1-f}{n} \; \frac{\sum_{i=1}^{n}[(y_i - x_i) - (\bar{y} - \bar{x})]^2}{n-1} = 0.5928$$

معامل التغير لتقدير الوسط الحسابي

$$CV(\bar{y}_d) = \frac{s(\bar{y}_d)}{\bar{y}_d} = \frac{\sqrt{0.5928}}{74.4} = 0.013$$

إن تقدير المجموع الكلي بطريق الفرق يعرف

$$Y_d = N\bar{y}_d$$

أما التباين وتقدير التباين إلى Y_d فهما على التوالي

$$var(Y_d) = N^2 \, var(\bar{y}_d)$$

$$s^2(Y_d) = N^2 \, s^2(\bar{y}_d)$$

8.7 التقديرات باستخدام الانحدار في العينة العشوائية البسيطة

يستخدم النموذج الخطي أو ما يسمى بخط الانحدار البسيط كثيراً في التحليل الإحصائي. وسوف نستخدمه هنا لتقدير الوسط الحسابي والمجموع للمجتمع. ويمكن التعبير عن العلاقة بين المتغيرين y وx بمعادلة خط الانحدار

$$y_i = \alpha + bx_i + \varepsilon_i$$

حيث إن $(0, \sigma^2)$ ε_i مهما كانت قيم x. وباستخدام النموذج الخطي فإننا نستخدم $\overline{y}_{lr}$ كتقدير إلى $\overline{Y}$ ، عندما تكون β ميل المستقيم معلوم يعرف $\overline{y}_{lr}$ كما يأتي

$$\overline{y}_{lr} = \overline{y} + \beta(\overline{X} - \overline{x})$$

عندما $\beta = 1$ فإن $\overline{y}_d = \overline{y}_{lr}$. واضح أن $\overline{y}_{lr}$ تقدير غير متحيز إلى $\overline{Y}$ مع تباين

$$\text{var}(\overline{y}_{lr}) = \frac{1-f}{n}(S_y^2 + \beta^2 S_x^2 - 2\beta S_{xy})$$

إن قيمة β التي سوف تعطينا أصغر تباين ممكن إلى $\overline{y}_{lr}$ هي

$$\beta_{opt} = \frac{S_{xy}}{S_x^2}$$

سيكون التباين الأمثل إلى $\overline{y}_{lr}$ كما يأتي

$$\text{var}(\overline{y}_{lr}) = \frac{1-f}{n} S_y^2 (1 - \rho^2)$$

حيث إن

$$\rho^2 = \frac{(S_{xy})^2}{S_x^2 S_y^2}$$

و ρ^2 يسمى معامل التحديد.

إن قيمة β تكون غير معلومة في أغلب الأحيان لذلك لا بد من تقديرها باستخدام البيانات التي حصلنا عليها من المسح. نستطيع تقدير قيمة β باستخدام ما يسمى بطريقة المربعات الصغرى

$$\beta = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - [\sum_{i=1}^{n} x_i]^2}$$

الآن نستخدم β بدلاً من β ليصبح على الوجه الآتي

$$\bar{y}_{lr} = \bar{y} + \beta\,(\bar{X} - \bar{x})$$

إن الشروط التي تصاحب النموذج الخطي قد لا تتحقق إذا كنا نسحب عينه من مجتمع محدود. ومن هذه الشروط يجب أن تكون العلاقة بين y و x خطية، كذلك تباينات y_i لقيمة محددة من x تبقى نفسها مهما كانت قيمة x لا بد من الإشارة هنا أنه إذا كان المجتمع صغيراً، وكذلك حجم العينة صغيراً، فإن تجاوز بعض شروط النموذج قد يكون حرجاً. وقد يخلق صغر حجم العينة بعض الصعوبات. ولكن إذا كان النموذج صحيحاً للبيانات فإن $\bar{y}_{lr}$ سيكون أفضل تقدير إلى $\bar{Y}$. وسيكون أفضل حتى من $\bar{y}_R$.

نستطيع أن نقدر تباين $\bar{y}_{lr}$ وذلك باستخدام طريقة المربعات الصغرى فإن

$$s^2(\bar{y}_{lr}) = \frac{1-f}{n}\,\frac{n-1}{n-2}\,(s_y^2 - \beta\, s_{xy})$$

أما المجموع الكلي للمجتمع فيمكن تقديره

$$Y_{lr} = N\bar{y}_{lr} = Y + \beta\,(X - X)$$

وكذلك فإن تقدير تباين يكون

$$s^2(Y_{lr}) = N^2\, s^2(\bar{y}_{lr})$$

مثال (6): للمثال (5) أوجد $\bar{y}_{lr}$ و $s(\bar{y}_{lr})$ وقارنها مع النتائج التي حصلت عليها إلى $\bar{y}$ و$\bar{y}_d$ و $\bar{y}_R$.

الحل:

$$\bar{y}_{lr} = \bar{y} + \beta(\bar{X} - \bar{x}) = 72.1 + 0.9925(74 - 71.7) = 74.383$$

$$s^2(\bar{y}_{lr}) = \frac{1-f}{n}\frac{n-1}{n-2}(s_y^2 - \beta s_{xy})$$

$$= \frac{1-(10/180)}{10}\frac{9}{8}[5981.48 - (0.9925)(6020.59)] = .642$$

الجدول الآتي يلخص النتائج التي حصلنا عليها

المقدر	التقدير	تقدير الخطاء المعياري
y بمفرده	$\bar{y} = 72.10$	$s(\bar{y}) = 23.767$
النسبة	$\bar{y}_R = 74.41$	$s(\bar{y}_R) = 0.833$
الفرق	$\bar{y}_d = 74.40$	$s(\bar{y}_d) = 0.770$
الانحدار	$\bar{y}_{lr} = 74.38$	$s(\bar{y}_{lr}) = 0.801$

إن المفاضلة بين $\bar{y}_d$ و$\bar{y}$ كتقديرين للوسط الحسابي للمجتمع يعتمد على ميل المستقيم (أي خط انحدار y على x (فإن $\bar{y}_d$ يفضل على $\bar{y}$ إذا كان ميل المستقيم قريباً من القيمة المثلى أي

$$\beta_{opt} = \frac{S_{xy}}{S_x^2}$$

فإذا كانت

$$\beta_{opt} > \frac{1}{2}$$

يكون

$$\text{var}(\bar{y}_d) < \text{Var}(\bar{y})$$

إذاً $\bar{y}_d$ يفضل على $\bar{y}$. من خلال هذه النتائج، واضح أن $\bar{y}_d$ يفضل على $\bar{y}_R$ و $\bar{y}_{lr}$ وجميعهم يفضلون على $\bar{y}$.

بالأسلوب نفسه نستطيع أن نقارن بين $\bar{y}_d$ و $\bar{y}_R$ لنصل إلى أن $\bar{y}_d$ يفضل على $\bar{y}_R$ إذا كانت

$$\frac{1}{2} < \beta < \frac{R+1}{2}$$

حيث إن β تمثل ميل المستقيم و R نسبة المجتمع. إذا لم تكن β و R معلومتين، نقدرهما من البيانات لنحصل على

$$\beta = \frac{s_{xy}}{s_x^2} \text{ و } R = \frac{\bar{y}}{\bar{x}}$$

إذاً $\bar{y}_d$ يفضل على $\bar{y}_R$ إذا كانت

$$\frac{1}{2} < \beta < \frac{R+1}{2}$$

نلاحظ من النتائج التي حصلنا عليها في الجدول السابق أن $\bar{y}$ يُعد أسوأ تقدير إلى $\bar{Y}$ حيث أعطانا أكبر تقدير للخطأ المعياري بينما التقديرات الأخرى متقاربة من حيث قيمة الخطأ المعياري، ممكن أن يكون أفضلهم $\bar{y}_d$ لهذا المثال.

9.7 التقديرات باستخدام الانحدار والفرق في العينة الطبقية

هناك طريقتان لتقدير الوسط الحسابي في العينة العشوائية الطبقية باستخدام خط الانحدار.

1. **الوسط الحسابي الطبقي المنفصل بطريقة الانحدار**

يعرف الوسط الحسابي بطريقة الانحدار للطبقة i

$$\bar{y}_{lri} = \bar{y}_i + \beta_i(\bar{X}_i - \bar{x}_i)$$

حيث إن β_i تقدير لميل المستقيم في الطبقة i و $\overline{X}_i$ يمثل الوسط الحسابي للطبقة. إذاً الوسط الحسابي الطبقي المنفصل بطريقة الانحدار يعرف

$$\overline{y}_{lrs} = \sum_{i=1}^{K} W_i \, \overline{y}_{lri} = \sum_{i=1}^{K} W_i [\overline{y}_i + \beta_i (\overline{X}_i - \overline{x}_i)]$$

أما المجموع الكلي الطبقي المنفصل بطريقة الانحدار

$$Y_{lrs} = N\overline{y}_{lrs}$$

أما تباين $\overline{y}_{lrs}$ يمكن إيجاده

$$\text{var}(\overline{y}_{rs}) = \sum_{i=1}^{K} W_i^2 Var(\overline{y}_{lri}) = \sum_{i=1}^{K} W_i^2 \frac{1-f_i}{n_i} (S_{yi}^2 + \beta_i^2 S_{xi}^2 - 2\beta_i S_{xyi})$$

ويمكن تقدير تباين بما يأتي

$$s^2(\overline{y}_{rs}) = \sum_{i=1}^{K} W_i^2 \frac{1-f_i}{n_i} (s_{yi}^2 + \beta_i^2 s_{xi}^2 - 2\beta_i s_{xyi})$$

2 الوسط الحسابي الطبقي المشترك بطريقة الانحدار

يعرف على الوجه الآتي

$$\overline{y}_{lrc} = \overline{y}_{st} + \beta(\overline{X} - \overline{x}_{st}) = \sum_{i=1}^{K} W_i [\overline{y}_i + \beta(\overline{X}_i - \overline{x}_i)]$$

أما المجموع الكلي الطبقي المشترك بطريقة الانحدار فيعرف

$$Y_{lrc} = N\overline{y}_{lrc}$$

أما تباين $\overline{y}_{lrc}$ فيمكن إيجاده

$$Var(\overline{y}_{lrc}) = \sum_{i=1}^{K} W_i^2 \frac{1-f_i}{n_i} (S_{yi}^2 + \beta^2 S_{xi}^2 - 2\beta S_{xyi})$$

حيث إن

$$\beta = \frac{\sum_{i=1}^{K} W_i^2 \frac{1-f_i}{n_i} S_{xyi}}{\sum_{i=1}^{K} W_i^2 \frac{1-f_i}{n_i} S_{xi}^2}$$

ويمكن تقدير التباين وذلك باستبدال β و S_{yi}^2 و S_{xi}^2 و S_{xyi} بتقديراتها من العينة وهي على التوالي β و s_{yi}^2 و s_{xi}^2 و s_{xyi} .

النقطة الأخيرة في هذا الفصل إذا عوضنا عن $\beta = 1$ في الوسط الحسابي الطبقي المشترك بطريقة الانحدار نحصل على

$$\bar{y}_{dst} = \bar{y}_{st} + (\bar{X} - \bar{x}_{st})$$

وهذا هو الوسط الحسابي الطبقي المشترك بطريقة الفرق مع تباين

$$Var(\bar{y}_{dst}) = \sum_{i=1}^{K} W_i^2 \frac{1-f_i}{n_i}(S_{yi}^2 + S_{xi}^2 - 2S_{xyi})$$

تمارين

1. لتحديد نسبة المصروف من دخل العائلة الشهري على الطعام، سحبت عينة عشوائية بسيطة حجمها n=12 عائلة من منطقة سكنية صغيرة، عدد العائلات فيها N=200 والنتائج كما هي مبينة في الجدول الآتي:

العائلة	1	2	3	4	5	6	7	8	9	10	11	12
الدخل	200	170	450	300	600	550	1000	1200	150	220	250	320
المصروف	125	110	200	170	250	240	350	500	105	130	140	160

أ- قدر النسبة R ثم أوجد تقدير الخطأ المعياري.

ب- أوجد فترة 95% ثقة للنسبة R.

2. استخدم البيانات في السؤال السابق علماً بأن معدل دخل العائلة الشهري في القرية $\bar{X} = 455$ أوجد:

أ. تقدير معدل مصروفات العائلة الشهرية على المواد الغذائية بالاعتماد على المعلومات المتوافرة حول مصروفات العائلة على المواد الغذائية وحدها.

ب. استخدم النسبة لتقدير معدل مصروفات العائلة الشهرية على المواد الغذائية.

ت. استخدم خط انحدار y على x لتقدير مصروفات العائلة الشهرية على المواد الغذائية.

ث. أوجد تقدير الخطأ المعياري للحالات الثلاثة أعلاه، ثم علق على النتائج التي حصلت عليها.

3. يقوم أحد الخبراء بتخمين وزن التفاح الموجود على كل شجره، إذا كان لدينا حديقة فيها 100 شجرة، خُمن وزن التفاح لهذه الأشجار 12500=x كيلو غرام، سحبت عينة عشوائية بسيطة حجمها 10 شجرات، وجرى قطف التفاح وجرى وزنه لإيجاد الوزن الحقيقي للتفاح في كل شجرة، الجدول التالي يبين الوزن التخميني والحقيقي لعشر شجرات

الشجرة	1	2	3	4	5	6	7	8	9	10
الوزن الحقيقي y_i	55	42	46	39	71	61	58	57	58	67
الوزن التخميني x_i	56	47	48	40	78	59	52	58	55	67

أ. قدر وزن التفاح الموجود في الحديقة باستخدام طريقة الفرق، ثم أوجد الانحراف المعياري.

ب. قدر وزن التفاح الموجود في الحديقة باستخدام خط انحدار y على x ثم أوجد الانحراف المعياري للتقدير.

ت. قارن بين الطريقتين، أيُّها في اعتقادك أفضل لتقدير وزن التفاح الموجود في الحديقة؟

4. استخدم البيانات السؤال الأول لتحديد حجم العينة اللازم لتقدير النسبة R إذا كان حد الخطأ لتقدير النسبة d=0.15.

5. استخدم البيانات المعطاة في السؤالين الأول والثاني لتحديد حجم العينة المطلوب، لتقدير الوسط الحسابي للطرق الثلاثة إذا كان مقدار حد الخطأ لتقدير الوسط الحسابي d=0.15.

6. .سحبت عينة عشوائية بسيطة بحجم 5 أشخاص من الذين أنهوا برنامجاً للحمية، الجدول الآتي يبين الأوزان قبل وبعد الالتحاق بالبرنامج.

الشخص	1	2	3	4	5
الوزن قبل البرنامج	120	140	160	130	180
الوزن بعد البرنامج	100	105	110	115	120

أ. قدر R وأوجد فترة 95% ثقة للقيمة الحقيقية للنسبةR.

ب. قدر معدل أوزان الأشخاص الملتحقين بالبرنامج بعد إتمامه ومن ثم أوجد فترة 90% ثقة باستخدام النسبة.

7. لنفترض أن عينة عشوائية بسيطة حجمها 10 طلاب اختيرت من طلبة أحد الأقسام العلمية في الجامعة، وسئلوا عن مقدار دخلهم الشهري (x) ومصروفاتهم الشهرية لشراء الكتب العلمية (y). الجدول الآتي يعطينا إجابات الطلاب. علماً بأن عدد طلاب القسم يبلغ 500 طالب.

الطالب	1	2	3	4	5	6	7	8	9	10
(x)	20	15	22	30	40	35	25	32	42	24
(y)	1100	1200	1300	1400	1600	1300	1250	1300	1650	1200

أ. قدر معدل مصروفات الطلاب في القسم على مشتريات الكتب باستخدام النسبة R ومقدار الخطأ المعياري للتقدير.

ب. قدر معدل مصروفات الطلاب الشهرية لشراء الكتب العلمية باستخدام خط انحدارy على x ومن ثم أوجد مقدار الخطأ المعياري للتقدير.

ت. استخدم y بمفرده لتقدير معدل مصروفات الطلاب الشهرية ولشراء الكتب العلمية ومن ثم أوجد مقدار الخطأ المعياري للتقدير.

ث. قارن بين ما حصلت عليه باستخدام الطرق الثلاث أعلاه ومن ثم أوصِ بأفضلهم لكي تستخدم لاحقاً في مثل هذه التقديرات.

8. ترغب الغرفة الصناعية في تقدير دخل الشركات الصناعية لعام 2005 العاملة ضمن المنطقة الموجودة فيها، يبلغ عدد الشركات الصناعية25 شركة. الدخل لجميع الشركات الصناعية لعام 2001 بلغ 670 مليون دولار. قامت الغرفة بسحب عينة عشوائية حجمها 9 شركات وحصلت على معلومات حول دخل هذه الشركات لعامي 2001 و2005، الجدول الآتي يعطينا الدخل بملايين الدولارات

الشركة	1	2	3	4	5	6	7	8	9
2001	20	19	30	26	25	33	14	39	27
2005	23	24	28	27	30	38	16	44	29

أ. استخدم النسبة لتقدير مجموع دخل الشركات الصناعية لعام 2005 ومن ثم أوجد فترة 95% ثقة للدخل الحقيقي للشركات.

ب. استخدم الفرق لتقدير مجموع دخل الشركات الصناعية لعام 2005 ومن ثم أوجد فترة 95% ثقة للدخل الحقيقي للشركات.

ت. استخدم طريقة الانحدار لتقدير مجموع دخل الشركات الصناعية لعام 2005 ومن ثم أوجد فترة 95% ثقة للدخل الحقيقي للشركات.

ث. أي الطرق الثلاثة الواردة أعلاه أفضل لتقدير مجموع دخل الشركات الصناعية؟ ولماذا؟

9. يوجد في إحدى المدن 500 شركة يعمل في هذه الشركات 500.000 موظف. جرى سحب عينة عشوائية بحجم 5 شركات وحصلنا على البيانات الآتية:

عدد الموظفين(X)	30	12	15	25	35
المصروف على الإنترنت بآلاف الدولارات (y)	10	12	15	25	35

استخدم طريق الانحدار لتقدير المصروفات الشهرية على الإنترنت ومن ثم أوجد فترة 90% ثقة

للقيمة الحقيقية للمصروفات الشهرية على الإنترنت.

10. يقوم أحد المصانع بإنتاج نوعين من السلع لغرض عرضهما في الأسواق. يمكن أن تعد النوعين من السلع، طبقتين لتقدير إجمالي حجم المبيعات من السلعتين. سحبت عينة عشوائية بسيطة من زبائن السلعتين، وسئلوا لإعطاء عدد الوحدات التي يمكن أن يشتروها من السلعتين للسنة القادمة y علماً بأن عدد الوحدات التي قام بشرائها نفس الزبائن من السلعتين للسنة الماضية x متوافرة لدينا، الجدول الآتي يعطينا البيانات للسلعتين. علماً بأن $X_1 = 24500$ مجموع الوحدات التي بيعت السنة الماضية من السلعة الأولى، وعدد الزبائن للسلعة الأولى $N_1 = 120$. كذلك فإن عدد الزبائن للسلعة الثانية للسنة الماضية ومجموع الوحدات التي بيعت من السلعة الثانية هي على التوالي $N_2 = 180$ و $X_2 = 21200$.

السلعة	x_i	204	143	82	256	198				
الأولى	y_i	210	160	75	280	190				
السلعة	x_i	137	189	119	63	103	107	159	63	87
الثانية	y_i	150	200	125	60	110	100	180	25	90

أ. قدر إجمالي المبيعات للسلعتين للسنة القادمة باستخدام طريقة النسبة وخط انحدار y على x والفرق.

ب. أوجد تقدير الخطأ المعياري للطرق الثلاثة أعلاه، ثم وازن بين هذه الطرق.

ت. ما هي الطريقة التي تفضل استخدامها لتقدير إجمالي المبيعات، ولماذا؟

المراجع العربية

1. سليمان محمد طشطوش. (2001) أساسيات المعاينة الإحصائية، دار الشروق للنشر والتوزيع، عمان.

2. حسين علوان (1994). طرق المعاينة، دار الفرقان، عمان.

3. هلال عبود البياتي وصبري رديف العاني (1981). العينات، جامعة بغداد، بغداد.

4. وليد عبدالحميد نوري وعبدالمجيد حمزة الناصر (1981). العينات، وزارة التعليم العالي والبحث العلمي، بغداد.

5. وليم كوكوران – ترجمة أنيس كنجو (1995). تقنية المعاينة الإحصائية، جامعة الملك سعود، الرياض.

References

1. Barnett, V. (1991). Sampling Survey Principles and Methods, Edward Arnold, London.
2. Benedetto, J. J. and Ferreira, P. J. (2001). Modern Sampling Theory, Birkhauser Boston.
3. Brewer, K. W. R. (1963). Ratio Estimation in Finite Populations: Some Results Deducible from the Assumption of an Underlying Stochastic Process, *Aust. J. Statistics*. 5, 93–105.
4. Cassel, C. M., Sarndal, C. E. and Wretman, J. H. (1976). Some Results on Generalized Difference and Generalized Regression Estimation for Finite Population, *Biometrika* 63, 615–620.
5. Chaudhuri, A. and Stenger, H. (1992). Survey Sampling: Theory and Methods, Marcel Dekker, New York.
6. Cochran, W. G. (1977). Sampling Technique, 3rd Ed. Wiley, New York.
7. David, I. P. and Sukhatme, B. V. (1974). On the Bias and Mean Square Error of the Ratio Estimator, *J. Amer. Statist. Asso.* 69, 464–466.
8. Deng, L. Y. and Wu, C. F. J. (1987). Estimation of the Variance of the Regression Estimator, *J. Amer. Statist. Asso.* 82, 568–576.
9. Durbin, J. (1959). A Note on the Application of Quenouille's Methods of Bias Reduction to Estimation of Ratios, *Biometrika* 46, 477–480.
10. Foreman, E. K. (1991). Survey Sampling Principles, Marcel Dekker, New York.
11. Goodman, L. A. and Hartley, H. O. (1958). The Precision of Unbiased Ratio-type Estimators. *J. Amer. Statist. Asso.* 53, 491–508.
12. Govindarajulu, Z. (1999). Elements of Sampling Theory and Methods, Prentice Hall.
13. Hajek, J (1981). Sampling from Finite Population, Marcel Dekker, New York.
14. Hansen, M. H., and Hurwitz, W. N. (1943). On the Theory of Sampling from Finite Populations. *Ann. Math. Statist.* 14,333–362.
15. Hedayat, A. S. and Sinha, B. K. (1991). Design and Inference in Finite Population Sampling, Wiley, New York.

16. Hartley, H. O. and Ross, A. (1954). Unbiased Ratio Estimates, *Nature* 174, 270–271.
17. Kish, L., Namboodiri, N. K. and Pillai, R. K. (1962). The Ration bias in Surveys. *J. Amer. Stat. Assoc.* 57, 863–876.
18. Kish, L (1995). Survey Sampling, Wiley, New York.
19. Lahiri, D. B. (1951). A Method for Sample Selection Providing Unbiased Ratio, Estimate. *Bull. Int. Stat. Inst.* 33, 133–140.
20. Levy, P. S. and Lemeshow S. (1991). Sampling of Population: Methods and Applications, Wiley, New York.
21. Lohr, S. L. (1999). Sampling: Design and Analysis, Duxbury, New York.
22. Mickey, M. R. (1959). Some Finite Population Unbiased Ratio and Regression Estimators, *J. Amer. Statist. Asso.* 54, 594–612.
23. Murthy, M. N. (1957). Ordered and Unordered Estimators in Sampling without Replacement. *Sankhya* 18, 379–390.
24. Nanjamma, N. S., Murthy, M. N. and Sethi, V. K. (1959). Some Sampling Systems Providing Unbiased Ratio Estimators, *Sankhya* 21, 299–314.
25. Olkin, I. (1958). Multivariate Ratio Estimation for Finite Populations, *Biometrika* 45, 154–165.
26. Quenouille, M. H. (1956). Notes on Bias in Estimation *Biometrika* 43, 353–360.
27. Raj, D. (1968). Sampling Theory, McGraw Hill, New York.
28. Rao, J. N. K. (1968). Some Small Sample Results in Ratio and Regression Estimation. *J. Ind. Stat. Assoc.*, 6, 160–168.
29. Rao, J. N. K. (1969). Ratio and Regression Estimators. In: New Developments in Survey Sampling (Johnson, N. L. and Smith, H. Jr. Eds.) 213–234, Wiley, New York.
30. Rao, J. N. K. and Vijayan, K. (1977). On Estimating the Variance in Sampling with Probability Proportion to Aggregate Size, *J. Amer. Statist. Asso.* 72, 579–584.
31. Rao, J. N. K. and Webster, J. T. (1966). On Two Methods of Biased Reduction in the Estimation of Ratio, *Biometrika* 53, 315–321.
32. Rao, J. N. K. (1988). Ratio and Regression Estimators. In: Handbook of Statistics, Vol. 6, Sampling (Krishnaiah, P. R. and Rao, C. R, Eds.), 449–468, North Holland, Amsterdam.
33. Rao, P. S. R. S. and Rao, J. N. K. (1971). Small Sample Results for Ratio Estimation, *Biometrika* 58, 625–630.
34. Rao, T. J. (1981 a). A Note on Unbiasedness in Ratio Estimators, *J. Statist. Plann. Infere.*, 5, 335–340.
35. Rao, T. J. (1981 b). On a Class of Almost Unbiased Ratio Estimators, *Ann. Inst. Statist. Math*, 33, 225–231.
36. Ray, S. K. and Sahai, A. (1980). Efficient Families of Ratio and Product Type Estimators, *Biometrika* 67, 211–215.
37. Royall, R. M. (1970). On Finite Population Sampling Theory under Certain Linear Regression Models, *Biometrika* 57, 377–387.
38. Royall, R. M. and Cumberland, W. G. (1981 a). An Empirical Study of the Ratio Estimator and Estimators of its Variance, *J. Amer. Statist. Asso.* 76, 66–77.
39. Royall, R. M. and Cumberland, W. G. (1981 b). The Finite Population Linear Regression Estimator and Estimators of its Variance: An Empirical Study, *J. Amer. Statist. Asso.* 76, 924–930.

40. Royall, R. M. and Eberhardt, K. R. (1975). Variance Estimation of the Ratio Estimator, *Sankhya* C 37, 43–52.
41. Sampath, S. (2000). Sampling Theory and Methods, CRC Press.
42. Singh, P. and Srivastava, A. K. (1980). Sampling Schemes Providing Unbiased Regression Estimators, *Biometrika* 67, 205–209.
43. Singh, R. and Sukhatme, B. V. (1973). Optimum Stratification with Ratio and Regression Methods of Estimation, *Ann. Inst. Statist. Math.* 25, 627–633.
44. Scheaffer, R.L., Mendenhall, W. and Ott L. (1996). Elementary Sampling Survey, 5th ed., Duxbury, New York.
45. Smith, T.M.F. (1976). The Foundation of Survey Sampling: A review, *Journal of Royal Statistics Society* A139, 183–204.
46. Srivenkataramana T. (1980). A Dual to Ratio Estimators in Sample Surveys, *Biometrika* 67, 199–204.
47. Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., and Ashok, C. (1984). Sampling Theory of Surveys with Applications, 3rd ed., Ames (Iowa): Iowa State University Press.
48. Thompson, S. K. (2002). Sampling, 2end Wiley, New York.
49. Tin, M. (1965). Comparison of some Ratio Estimators, *J. Amer. Statist. Asso.* 60, 294–307.
50. Tryfos, P. (1996). Sampling Methods for Applied Research, Wiley, New York.
51. Williams, W. H. (1963). The Precision of some Unbiased Regression Estimators, *Biometrics* 19, 352–361.
52. Yates, F. (1981). Sampling Methods for Censuses and Surveys 4th Ed., Grittin, London

الفصل الثامن

العينة العشوائية المنتظمة
Systematic Random Sample

1.8 مقدمة

لنفرض أننا نرغب في سحب عينة حجمها n من مجتمع يحتوي على N من الوحدات، ويوجد لدينا كشف كامل بأسماء هذه الوحدات، يمكن أن نحدد أعضاء العينة في الكشف ومن ثم نقوم بمشاهدة هذه الوحدات في المجتمع، لجمع المعلومات المطلوبة منها. فمثلاً يمكن أن نختار مجموعة من طلاب الجامعة، وذلك باستخدام الكشوفات التي تنشرها الجامعة بأسماء طلابها، أو يمكن أن نختار مجموعة من الكتب، وذلك باستخدام الكشوفات الموجودة في المكتبة بأسماء الكتب المحفوظة فيها. إن اختيار عينة عشوائية بسيطة يبدو صعباً جداً في مثل هذه الحالات، ويحتاج إلى وقت وجهد كبيرين. لنتصور أننا نرغب في اختيار 500 طالب من بين 15000 طالب مسجلين في إحدى الجامعات، لذلك هناك رغبة كبيرة للبحث عن طريقة بديلة.

إن الطريقة الشائعة الاستخدام في مثل هذه الحالات هي ما تسمى بالعينة العشوائية المنتظمة وهذه تسمى أحيانا العينة المنتظمة. ونستطيع أن نعرف العينة العشوائية المنتظمة على أنها العينة التي يتم اختيارها بواسطة اختيار وحدة واحدة بطريقة عشوائية من بين أول k من العناصر الموجودة في الكشف

أو الإطار، ثم بعد ذلك تؤخذ بقية الوحدات بشكل تكون فيه الوحدات المتتالية ذات أبعاد متساوية فيما بينها والعدد منها مساوٍ إلى k.

لنَعد الآن إلى مثالنا الذي أردنا فيه اختيار 500 طالب من بين 15000 من طلبة الجامعة. لا بد من تحديد قيمة k الذي يمكن أن نجد قيمتها باستخدام $k = N/n$ ففي مثالنا هذا $k = 30$ بطبيعة الحال فإن الكشف بأسماء الطلاب مرقم من 1 إلى 15000 وإذا لم يكن مرقماً بهذه الطريقة فإنه من السهولة القيام بترقيمه. نقوم الآن باختيار رقم وبطريقة عشوائية من بين أول ثلاثين رقماً أي بين 1 إلى30. ولنفرض أننا اخترنا الرقم 7 بطريقة عشوائية فسيكون الطالب رقم 7 في العينة، ومن ثم نقوم بأخذ الطالب من المجموعة الثانية أي من بين الطلاب الذين أرقامهم بين 31 إلى 60 والذي يكون تسلسله $7 + k = 37$ ومن المجموعة الثالثة الطالب الذي تسلسله $7 + 2k = 67$ وهكذا إلى نهاية الكشف.

2.8 طرق اختيار وحدات العينة المنتظمة

نستخدم العينة العشوائية المنتظمة باستمرار، إذا كان هناك إطار كامل وحديث متوافر. سوف نتكلم عن طريقتين لاختيار وحدات العينة المنتظمة من المجتمع.

1.2.8 الطريقة الخطية المنتظمة

إن هذه الطريقة هي الطريقة الشائعة الاستخدام. نفرض أن لدينا مجتمعاً مكوناً من N عنصر، وهذه العناصر مرتبة بطريقة خطية، تعطى هذه العناصر أرقاماً من 1 إلى N ثم نجد قيمة $k = N/n$، ونقوم باختيار عنصر من بين 1 إلى k بطريقة عشوائية ولنفرض أننا اخترنا العنصر رقم i حيث إن i أقل أو يساوي k، تسمى k بفترة الاختيار. وسوف تحتوي العينة التي تسحب بهذه

الطريقة على العناصر $i, i+k, i+2k, ..., i+(n-1)k$ باستخدام هذه الطريقة سوف نحصل على k من العينات الممكنة السحب. إن الجدول الآتي يعطينا العينات الممكنة السحب حيث إن y_i ترمز إلى قيمة العنصر i.

رقم العينة

1	2	...	i	...	k
y_1	y_2	...	y_i	...	y_k
y_{1+k}	y_{2+2k}	...	y_{i+2k}	...	y_{2k}
⋮	⋮	⋮	⋮	⋮	⋮
$y_{1+(n-1)k}$	$y_{2+(n-1)k}$	...	$y_{i+(n-1)k}$	...	y_{nk}

هناك مشكلة تواجهنا عندما لا تكون $k=N/n$ مساوية إلى عدد صحيح، نقوم بتقريب k إلى أقرب عدد صحيح. إن هذه الطريقة يمكن أن تؤدي لأن تكون العينات الممكنة السحب من المجتمع غير متساوية الحجم.

مثال (1): نفرض أن لدينا مجتمعاً مكوناً $N=11$ وحدة وأردنا أن نسحب عينة منتظمة حجمها $n = 4$ في هذه الحالة $k=11/4=3$.سيكون لدينا العينات الممكنة الظهور هي (1,4,7,10) أو (2,5.8,11) أو(3,6,9). واضح أن العينة الأخيرة عدد وحداتها 3 وهو لا يساوي عدد وحدات العينات الأولى.

2.2.8 طريقة الاختيار الدائرية المنتظمة

لحل مشكلة عدم تساوي عدد الوحدات في العينات الممكنة الظهور عندما تكون $N \neq nk$ يمكن استخدام طريقة الاختيار الدائرية. إن الطريقة الجديدة تتضمن اختيار وحدة بطريقة عشوائية من بين العناصر 1 إلى N ولنفرض أن هذا العنصر رقمه i و$i \leq N$ ثم نقوم باختيار العناصر $i+jk$ إذا كانت $i+jk \leq N$ لكل $j=0,1,2,\dots,(n-1)$ وكذلك نختار العناصر

$i+jk$□N إذا كانت $i+jk>N$ إن قيمة k تكون عدداً صحيحاً أو مقرباً إلى أقرب عدد صحيح.

مثال (2): لنَعُد إلى مثالنا السابق الذي فيه $N=11$ و$n=4$ و$k=3$ لنستخدم الطريقة الدائرية لاختيار جميع العينات الممكنة.

وهي كما يأتي: (4,7,10,2) و(3,6,9,1) و(2,5,8,11) و(1,4,7,10) و(9,1,4,7) و(8,11,3,6) و(7,10,2,5) و(6,9,1,4) و(5.8.11.3) و(10.2.5.8) و(11,3,6,9).

يمكن بسهولة أن نثبت أن كل وحدة من وحدات المجتمع لها نفس الفرصة في الظهور. وكذلك واضح أن كل عينة من العينات الممكنة الظهور وبحجم n أيضاً لها الفرصة نفسها في الظهور.

3.8 مزايا وعيوب العينة المنتظمة

إن من أهم مزايا العينة المنتظمة

1. أنها سهلة الاستخدام أو التنفيذ في الميدان، ولذلك تُعد أقل عرضة لأن يخطئ العاملون في اختيار وحداتها بالمقارنة إلى العينة العشوائية البسيطة أو العينة العشوائية الطبقية خصوصاً إذا لم يكن هناك إطار جيد متوافر.
2. أنها توفر معلومات أكثر لكل وحدة تكلفة بالقياس إلى العينة العشوائية البسيطة.
3. أنها تغطي المجتمع بصورة أفضل من العينة العشوائية البسيطة أو الطبقية. لذلك نراها تستخدم كثيراً في الغابات لتقدير كمية الخشب الموجودة وفي البحيرات لتقدير عدد الأسماك.
4. يمكن استخدام العينة العشوائية المنتظمة حتى ولو كنا لا نعرف قيمةN.

5. أحيانا تعطي تقديراً أكثر دقة من العينة العشوائية البسيطة، خصوصاً إذا كان المجتمع لا يحتوي على دورات معينة أو مرتباً ترتيباً تصاعدياً أو تنازلياً.

أما أهم عيوبها فهي

1. إذا كان المجتمع يحتوي على دورات متكررة فيجب استخدام العينة المنتظمة بحذر شديد.
2. لا يمكن الحصول على تقدير غير متحيز لتباين تقدير الوسط الحسابي أو المجموع الكلي باستخدام عينة واحدة، لأننا استخدمنا العشوائية في الاختيار مرة واحدة، أي باختيار عنصر واحد فقط من عناصر العينة.
3. إذا كان المجتمع يحتوي على دورات غير معروفة لنا فإن استخدام العينة المنتظمة يؤدي إلى تحيز في تقدير الوسط الحسابي أو المجموع.

4.8 تقدير الوسط الحسابي والمجموع الكلي للمجتمع

إن الهدف الرئيس لأكثر المسوحات بالعينة هو تقدير مَعْلَم أو أكثر من مَعَالم المجتمع. لذلك سوف نتكلم عن تقدير الوسط الحسابي والمجموع للمجتمع.

1.4.8 تقدير الوسط الحسابي للمجتمع

يمكننا تقدير الوسط الحسابي للمجتمع باستخدام العينة العشوائية المنتظمة باستخدام الوسط الحسابي للعينة

$$\bar{y}_{sy} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

كذلك يمكن تقدير تباين الوسط الحسابي للعينة $\bar{y}_{sy}$ بما يأتي:

$$s^2(\bar{y}_{sy}) = \frac{N-n}{N}\frac{s^2}{n} \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y}_{sy})^2$$

إذا كانت N غير معلومة فإننا نهمل معامل التصحيح للمجتمع المحدود $\frac{N-n}{N}$. إن تقدير الوسط الحسابي للمجتمع باستخدام $\bar{y}_{sy}$ الوسط الحسابي للعينة المنتظمة، وكذلك تقدير التباين للوسط الحسابي للعينة $\bar{y}_{sy}$ مماثل إلى الوسط الحسابي وتقدير تباين الوسط الحسابي للعينة العشوائية البسيطة $\bar{y}$ ولكن هذا لا يعني أن تباين المجتمع للتقديرين متساويان حيث إن تباين $\bar{y}$

$$Var(\bar{y}) = \frac{\sigma^2}{n}\frac{N-n}{N-1} = \frac{S^2}{n}\left(\frac{N-n}{N}\right)$$

وتباين $\bar{y}_{sy}$

$$Var(\bar{y}_{sy}) = \frac{\sigma^2}{n}[1 + (n-1)\rho]$$

حيث إن ρ يمثل معامل الارتباط بين أزواج من وحدات العينة نفسها. واضح أن $Var(\bar{y})$ و $Var(\bar{y}_{sy})$ غير متساويين.

إذا كانت قيمة ρ قريبة من واحد هذا يعني أن العناصر داخل العينة المنتظمة متماثلة إلى حد كبير بالنسبة إلى المتغير الذي نحن بصدد دراسته، وهذا سوف يؤدي إلى أن يكون تباين الوسط الحسابي للعينة المنتظمة أكبر من تباين الوسط الحسابي للعينة العشوائية البسيطة. أما إذا كانت قيمة ρ سالبة فربما تكون العينة المنتظمة أفضل من العينة العشوائية البسيطة. نلاحظ أن قيمة ρ لا يمكن أن تكون كبيرة في السالب وإلا سيؤدي ذلك إلى أن تكون قيمة $Var(\bar{y}_{sy})$ سالبة.

لا نستطيع الحصول على تقدير غير متحيز إلى $var(\bar{y}_{sy})$ وذلك باستخدام عينة منتظمة واحدة، هذا لا يعني أنه لا يمكن الحصول على تقدير تباين $var(\bar{y}_{sy})$. عندما تكون العينة المنتظمة قريبة أو مماثلة إلى العينة البسيطة يمكن أن يكون تقدير تباين الوسط الحسابي للعينة المنتظمة $s^2(\bar{y}_{sy})$ قريباً جداً من تقدير تباين الوسط الحسابي للعينة العشوائية البسيطة $s^2(\bar{y})$.

2.4.8 تقدير المجموع الكلي للمجتمع

إن تقدير المجموع الكلي للمجتمع يتطلب معرفة حجم المجتمع N .ولكن في كثير من الحالات تكون N غير معروفة. لذلك كان أحد أسباب استخدام العينة العشوائية المنتظمة هو عدم معرفةN . ولكن إذا كانت N معروفة نستطيع أن نقدر المجموع الكلي للمجتمع باستخدام

$$Y_{sy} = N\bar{y}_{sy}$$

مع تقدير التباين

$$s^2(Y_{sy}) = N^2 s^2(\bar{y}_{sy}) = N^2 \frac{N-n}{N} \frac{s^2}{n}$$

نلاحظ أن هذه النتائج مماثلة لتلك التي حصلنا عليها باستخدام العينة العشوائية البسيطة لتقدير المجموع الكلي للمجتمع.

5.8 مقارنة العينة المنتظمة بالعينة العشوائية البسيطة

إن مقارنة العينة العشوائية المنتظمة بالعينة العشوائية البسيطة يعتمد إلى حد كبير على طبيعة المجتمع تحت الدراسة؛ لذلك سوف نقارن بين طريقتي المعاينة لأنواع محددة من المجتمعات.

1.5.8 المجتمع عشوائي (Random Population)

يمكن أن نقول إن المجتمع عشوائي إذا كانت عناصره لا تتبع ترتيباً معيناً أي مرتبة عشوائياً. إن وحدات العينة العشوائية المنتظمة التي تُسحب من مجتمع عشوائي يتوقع أن تكون مختلفة فيما بينها، أي إن $\rho \approx 0$ ؛ لذلك إذا كانت N كبيرة فإن $Var(\bar{y}_{sy}) \approx Var(\bar{y})$ أي تباين الوسط الحسابي للعينة المنتظمة مساوٍ تقريباً إلى تباين الوسط الحسابي للعينة العشوائية البسيطة. فمثلاً إذا أردنا أن نقدر المعدل التراكمي لطلبة الجامعة، فإن الإطار عبارة عن كشف بأسماء جميع طلبة الجامعة مرتبين حسب الأحرف الهجائية للاسم الأول للطالب أو الرقم الجامعي. واضح أنه لا توجد علاقة بين ترتيب أسماء الطلاب حسب الأحرف الهجائية أو أرقامهم الجامعية ومعدلاتهم التراكمية، لذلك نستطيع أن نَعُدَّ المجتمع عشوائيًا بالنسبة إلى علامات الطلاب، وفي هذه الحالة تكون العينة العشوائية المنتظمة مماثلة للعينة العشوائية البسيطة.

2.5.8 المجتمع مرتب (Ordered Population)

إن المجتمع يعتبر مرتباً إذا كانت عناصره مرتبة بناءً على مقادير بحسب مخطط معين. فمثلاً في حالات المسح لتقييم تدريس أحد المدرسين في مساق عدد الطلبة المسجلين فيه كبير، في العادة يطلب من الطلبة تقييم مدرسهم وفق نظام عددي من 1 إلى 10 حيث يكون واحد أقل علامة وعشرة أكبر علامة، ومن ثم يقومون بترتيب استمارات الإجابة وفق نظام معين، كأن نبدأ بأعلى العلامات ثم نبدأ في النزول حتى الوصول إلى أقل العلامات، ومن ثم نسحب عينة منتظمة من بين هذه الاستمارات، في هذه الحالة يكون المجتمع مرتبًا ترتيباً تنازلياً.

إن العينة المنتظمة التي تسحب من مجتمع مرتب بصورة عامة تكون عناصرها مختلفة، ومن ثم ستكون قيمة $\rho \leq 0$. ويمكن إثبات أنه إذا كانت N كبيرة وقيمة ρ أقل أو تساوي صفراً فإن $Var(\bar{y}_{sy}) \leq Var(\bar{y})$. هذا يعني أن العينة العشوائية المنتظمة المسحوبة من مجتمع مرتب تعطينا معلومات أكثر لكل وحدة تكلفة من العينة العشوائية البسيطة. نستطيع تقدير $Var(\bar{y}_{sy})$ باستخدام

$$s^2(\bar{y}_{sy}) = \frac{N-n}{N}\frac{s^2}{n}$$

في هذه الحالة يُعد التقدير محافظاً؛ لأنه سوف يكون أكبر مما يفترض أن يكون.

3.5.8 المجتمع دوري (Periodic Population)

يُعد المجتمع دورياً إذا كانت تغيرات عناصره تأخذ شكلاً دورياً. مثلاً إذا أردنا تقدير عدد المصلين الذين يرتادون مسجداً معيناً خلال شهر. بالتأكيد إذا أخذنا عدد المصلين الذين يأتون إلى صلاة الفجر يومياً مدة شهر سوف يكون تقديرنا لعدد المصلين أقل مما يجب، لأن عدد المصلين الذين يحضرون صلاة الفجر أقل من عددهم في الصلوات الأخرى. وبالمقابل لو أننا أخذنا عدد المصلين الذي يحضرون صلاة الظهر سوف يكون تقديرنا لعدد المصلين أكثر مما يجب، حيث إن عدد المصلين الذين يحضرون صلاة الظهر في المسجد يكون أكبر من بقية الصلوات وخصوصاً الذين يحضرون صلاة الجمعة. لذلك يُعد هذا النوع من المجتمعات مجتمعاً دورياً.

إن عناصر العينة المنتظمة المسحوبة من مجتمع دوري يمكن أن تكون متماثلة أو قريبة من بعضها البعض، ومن ثم سوف تكون قيمة ρ أكبر من صفر ($\rho > 0$). على سبيل المثال إن عدد المصلين الذين يصلون صلاة الفجر في

المسجد ستكون قريبة من بعضها البعض خلال أيام الشهر. يمكن أن نثبت إذا كانت N كبيرة و $\rho > 0$ فإن $Var(\bar{y}_{sy}) > Var(\bar{y})$. لذلك فإن العينة المنتظمة تعطينا معلومات أقل لكل وحدة كلفة بالمقارنة بالعينة العشوائية البسيطة إذا كان المجتمع دورياً.

للتخلص من مشكلة المجتمع الدوري يمكن أن نغير نقطة البداية العشوائية عدة مرات. إن هذه الطريقة سوف تقلل فرصة اختيار وحدات من المجتمع عندها نفس شكل التغير في المجتمع. فمثلاً في مثالنا أعلاه نستطيع أن نختار نقطة البداية بطريقة عشوائية، أي نختار صلاة من بين الصلوات الخمس خلال اليوم الأول من الأسبوع الأول، فمثلاً ظهرت صلاة المغرب نستمر مدة أسبوع بجمع بيانات حول عدد المصلين الذين يحضرون صلاة المغرب في المسجد حتى نهاية الأسبوع الأول، ومن ثم نختار وبطريقة عشوائية مرة أخرى إحدى الصلوات من بداية الأسبوع الثاني، ونجمع بيانات حول عدد المصلين الذين يصلون هذه الصلاة في المسجد مدة أسبوع آخر، وهكذا إلى نهاية الشهر.

يمكننا أن نفترض أن العينة المنتظمة التي تسحب بهذه الطريقة تكون مماثلة للعينة المنتظمة التي تسحب من مجتمع عشوائي، ومن ثم نستطيع أن نقدر تباين $\bar{y}_{sy}$ باستخدام

$$s^2(\bar{y}_{sy}) = \frac{N\square n}{N}\frac{s^2}{n}$$

6.8 تقدير النسبة

غالباً ما نرغب في استخدام البيانات التي حصلنا عليها باستخدام العينة العشوائية المنتظمة لتقدير النسبة في المجتمع P ويمكن الحصول عليها إذا عرفنا y_i على الوجه الآتي

$y_i = 1$ إذا كانت الوحدة تحمل صفة مميزة

$y_i = 0$ خلاف ذلك أي إن الوحدة لا تحمل الصفة المميزة

فيكون $\hat{p}_{sy}$ تقدير النسبة في العينة العشوائية المنتظمة

$$\hat{p}_{sy} = \bar{y}_{sy} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{a}{n}$$

حيث إن a عدد الوحدات في العينة الذين يحملون الصفة المميزة. أما تقدير التباين فيكون

$$s^2(\hat{p}_{sy}) = \frac{N-n}{N}\,\frac{\hat{p}_{sy}\hat{q}_{sy}}{n-1} = (1-f)\,\frac{\hat{p}_{sy}\hat{q}_{sy}}{n-1}$$

يمكن أن نهمل $(1-f)$ معامل التصحيح للمجتمعات المحدودة إذا كانت N كبيرة أو غير معلومة. نلاحظ هنا أن تقدير النسبة في العينة العشوائية المنتظمة مماثل لتقدير النسبة في العينة العشوائية البسيطة التي سبق أن تكلمنا عنها. لكن هذا لا يعني أن تباين المجتمع $Var(\hat{p}_{sy})$ للعينة المنتظمة مساوٍ $Var(\hat{p})$ للعينة العشوائية البسيطة. ومع ذلك إذا كانت N كبيرة وكذلك البيانات في العينة المنتظمة غير مرتبطة، أي $\rho = 0$ سيكون تباين المجتمعات للنسبتين $\hat{p}$ و $\hat{p}_{sy}$ متقاربين.

7.8 اختيار حجم العينة

إن تحديد حجم العينة من الأمور المهمة جداً في المسح بالعينة، وذلك لارتباط حجم العينة بالدقة ودرجة الثقة المطلوبتين للتقدير. ناهيك عن علاقة حجم العينة بتكاليف المسح بالعينة وسرعة تنفيذه.

1.7.8 اختيار حجم العينة لتقدير الوسط الحسابي

لتحديد حجم العينة المطلوب لتقدير الوسط الحسابي للمجتمع $\overline{Y}$ سوف نتبع الطريقة نفسها التي سبق أن اتبعناها في الفصول السابقة. لنفرض أننا حددنا مقدار الخطأ في التقدير بما يساوي d ومقدار الثقة بما يساوي 95% أي

$$P(\left|\overline{Y} - \overline{y}_{sy}\right| < d) = 0.95$$

فإن حجم العينة المطلوب يمكن إيجاده بحل المعادلة الآتي لإيجاد قيم n

$$2\sqrt{Var(\overline{y}_{sy})} = d$$

هذه المعادلة سوف تعتمد على قيمة S^2 و ρ اللذين يجب أن يكونا معلومين، على الأقل، القيمة التقريبية لهما، ويمكن إيجاد القيمة التقريبية باستخدام معلومات سابقه متوافرة من مسوحات مماثلة أو بسحب عينة أولية عشوائية من المجتمع. سوف لا نستخدم هذه المعلومات لإيجاد قيمة S^2 و ρ ولكن سوف نستخدم حجم العينة التي يمكن إيجادها في حالة العينة العشوائية البسيطة. وهذه الطريقة سوف تعطينا حجم عينة أكبر إذا كان المجتمع مرتباً وحجم عينة أقل إذا كان المجتمع دورياً؛ لاحظنا أن تباين الوسط الحسابي للعينة العشوائية البسيطة والوسط الحسابي للعينة العشوائية المنتظمة متقاربان إذا كان المجتمع عشوائياً. لذلك فإن حجم العينة المطلوب سيكون

$$n \geq \frac{NS^2}{ND + S^2}$$

حيث إن $D = \frac{d^2}{4}$ و S^2 على الأرجح تكون غير معلومة نحددها باستخدام معلومات سابقة أو سحب عينة عشوائية أولية.

2.7.8 اختيار حجم العينة لتقدير المجموع الكلي

سوف نتبع الطريقة نفسها التي اتبعناها في العينة العشوائية البسيطة لتحديد حجم العينة لتقدير المجموع الكلي للمجتمع Y وذلك باختيار d ليكون حد الخطأ لتقدير المجموع الكلي ودرجة الثقة لتكون 95%؛ لذلك فإننا نجد قيمة n بحل المعادلة

$$2\sqrt{N^2 Var(\bar{y}_{sy})} = d$$

لذلك فإن

$$n \ge \frac{NS^2}{ND+S^2}$$

حيث إن $D=\frac{d^2}{4N^2}$ ، ونقدر S^2 إذا كانت غير معلومة باستخدام معلومات سابقة أو سحب عينة عشوائية أولية.

مثال (3): إذا كانت شركة توزيع الكهرباء على المستهلكين في إحدى المدن ترغب في تقدير معدل وقت تأخير دفع فواتير الكهرباء من قبل المستهلكين. إن العينة العشوائية المنتظمة تبدو ملائمة لسحب عينة من بين 2500 فاتورة متأخرة الدفع. لقد أجرت البحث نفسه في السنة الماضية فوجدت أن $s^2 = 100$. ما هو حجم العينة الذي يلزمنا لتقدير الوسط الحسابي للمجتمع $\bar{Y}$ الذي يمثل معدل وقت تأخر دفع الفواتير. علماً بأنه تم تحديد حد الخطأ في التقدير $d = 2$ يوم.

الحل:

لابد من إيجاد قيمة

$$D=\frac{d^2}{4}=\frac{(2)^2}{4}=1$$

و

$$n \geq \frac{NS^2}{ND+S^2} \approx \frac{Ns^2}{ND+s^2} = \frac{(2500)(100)}{2500+100} = 96.19$$

لذلك لا بد لإدارة الشركة أن تسحب عينة حجمها $n = 97$ لتقدير معدل وقت التأخير في دفع الفواتير المتأخرة الدفع.

3.7.8 اختيار حجم العينة لتقدير النسبة

كذلك فإن الطريقة التي سوف تستخدم هنا لاختيار حجم العينة لتقدير النسبة في العينة العشوائية المنتظمة هي نفسها المستخدمة في العينة العشوائية البسيطة؛ ولذلك إذا حددنا حد الخطأ في التقدير بمقدار d ومقدار الثقة بمقدار 95% أي

$$P(|P - \hat{p}_{sy} < d|) = 0.95$$

نقوم بحل المعادلة

$$2\sqrt{Var(\hat{p}_{sy})} = d$$

لإيجاد قيمة n

$$n \geq \frac{NPQ}{ND+PQ}$$

حيث إن $D = \frac{d^2}{4}$ و $Q = 1 - P$. ولا بد من الإشارة هنا إلى أن قيمة P تكون غير معلومة ولا بد من تقديرها باستخدام معلومات سابقة أو سحب عينة عشوائية أولية. ويمكن استخدام $P=0.5$ في كل الأحوال، حيث إنه إذا استخدمنا $P=0.5$ فسوف يعطينا أكبر حجم عينة ممكناً وهذا يجعلنا

مطمئنين إلى أن حجم العينة الذي حصلنا عليه هو أكبر حجم يمكن الحصول عليه.

مثال (4) لمعرفة رأي سكان مدينة معينة في إنتاج إحدى الشركات المحلية فيما إذا كان مقبولاً أم لا. قرر الباحث سحب عينة منتظمة من سجلات الأحوال المدنية في المدينة، حيث هنالك P=5000 اسم لشخص عمره فوق 16 سنة مسجل في هذه السجلات. لقد تم تحديد مقدار الخطأ في تقدير النسبة للمجتمع P بما يساوي %3 أي d=0.03 . يبدو أنه لا توجد هنالك معلومات عن قيمة P لذلك قرر الباحث استخدام P=0.5 .

الحل:

لابد من إيجاد قيمة

$$D=\frac{d^2}{4}=\frac{(0.03)^2}{4}=0.000225$$

و

$$n\ {}^{3}\ \frac{NPQ}{ND+PQ}=\frac{(5000)(0.5)(0.5)}{5000(0.000225)+(0.5)(0.5)}=909.01$$

لا بدمن الإشارة هنا إلى أن قيمة n=910 هو أكبر قيمة ممكنة لحجم العينة حيث إن P=0.5 تعطينا هذه النتيجة كما أشرنا إلى ذلك سابقاً.

8.8 تكرار العينة المنتظمة

لقد سبق أن قلنا: إننا لا نستطيع أن نقدر تباين الوسط الحسابي للعينة المنتظمة $\bar{y}_{sy}$ باستخدام عينة واحدة. لذلك استخدمنا العينة العشوائية البسيطة لتقدير تباين $\bar{y}_{sy}$. ولكن في كثير من الحالات العملية، العينة المنتظمة لا تماثل العينة العشوائية البسيطة. إذاً لا بد من استخدام طريقة بديلة لتقدير تباين $\bar{y}_{sy}$. إن تكرار العينة العشوائية المنتظمة هو أحد الطرق التي يمكن

استخدامها. وكما هو واضح من التسمية لا بد من تكرار العينة المنتظمة للحصول على أكثر من عينة واحدة. فعلى سبيل المثال في مجتمع يحتوي على $N=250$ وحدة، نستطيع أن نحصل على 10 عينات حجم كل عينة 5 وحدات بسحب وحدة واحدة من بين كل 50 وحدة من وحدات المجتمع، ونرمز لها 10 1 in 50 وهذا يساوي عدد الوحدات لعينة واحدة إذا سحبنا وحدة واحدة من بين كل 5 وحدات، ونرمز لها 1 1 in 5.

نلاحظ أن الطريقتين سوف تعطينا عدد الوحدات نفسها التي سوف نسحبها من المجتمع لتقدير الوسط الحسابي للمجتمع $\overline{Y}$. ولكن الطريقة الثانية سوف تعطينا $n_s=10$ عينات نستطيع أن نستخدمها لتقدير $\hat{V}ar(\overline{y}_{sy})$.

لتوضيح فكرة اختيار العينات العشوائية المنتظمة المكررة من المجتمع لنفرض أن لدينا مجتمعاً حجمه $N=960$ وحدة والتي يمكن أن نعطيها أرقاماً من 1 إلى 960. إذا أردنا اختيار عينة منتظمة حجمها $n=60$ نختار $k=N/n=16$ ومن ثم نختار عدداً عشوائياً بين 1 إلى 16 ليكون نقطة البداية العشوائية للعينة، ونختار بقية الوحدات بالأسلوب الذي تكلمنا عنه سابقاً. الآن ما هي الطريقة التي يمكن استخدامها لاختيار $n_s=10$ عينات منتظمة بدلاً من العينة العشوائية أعلاه. نبدأ بتحديد قيمة $k'=n_s k$، لقد سبق أن حدنا $k=16$ إذاً $k'=10(16)=160$. إن الخطوة الثانية التي نقوم بها هي اختيار 10 أرقام وبطريقة عشوائية من بين الأرقام الواقعة بين 1 إلى 160 ليكون كل رقم من هذه الأرقام نقطة البداية العشوائية للعينات المنتظمة $n_s=10$. ومن ثم نكمل كل عينة بطريقة منتظمة كما مر معنا سابقاً، ولكن $k'=160$ لنحصل على 10 عينات حجم كل عينة 6 وحدات.

لنفرض أننا سحبنا 10 أرقام بطريقة عشوائية بين 1 إلى 160 وحصلنا على الأرقام التالية 66,18, 6, 22, 50, 70,140,120, 90,155 نكمل كل العينات لنحصل على الجدول 1.

جدول (1) اختيار 10 عينات منتظمة مكررة

العينة	نقطة البداية العشوائية	العنصر الثاني	العنصر الثالث	العنصر الرابع	العنصر الخامس	العنصر السادس
1	6	166	326	486	646	806
2	18	178	338	498	658	818
3	22	182	342	502	662	822
4	50	210	370	530	690	850
5	66	226	386	546	706	866
6	70	230	390	550	710	870
7	90	250	410	570	730	890
8	120	280	440	600	760	920
9	140	300	460	620	780	940
10	155	315	475	635	795	955

لقد استخدمنا $n_s = 10$ وذلك للحصول على عدد كافٍ من الأوساط الحسابية للعينات وهذا هو القدر الذي نحتاجه لتقدير $var(\bar{y}_{sy})$. لقد اخترنا $k' = 160$ لتعطينا العدد نفسه من العناصر الذي حصلنا عليه باختيار عينة منتظمة واحدة $1 \le in \le k$ و $k = 16$ وأخيراً فإن $k' = kn_s = 160$.

لتقدير الوسط الحسابي للمجتمع $\bar{Y}$ باستخدام $(1 \le in \le k)$ n_s عينة منتظمة سوف نستخدم

$$\bar{y}_{syw} = \frac{1}{n_s} \sum_{i=1}^{n_s} \bar{y}_i$$

حيث إن $\bar{y}_i$ يمثل الوسط الحسابي للعينة المنتظمة i و $i = 1, 2, ..., n_s$ وتقدير تباين $\bar{y}_{syw}$

$$s^2(\bar{y}_{syw}) = \frac{N - n}{N} \frac{\sum_{i=1}^{n_s} (\bar{y}_i - \bar{y}_{syw})^2}{n_s(n_s - 1)}$$

وكذلك يمكن استخدام العينات العشوائية المنتظمة المكررة لتقدير المجموع الكلي للمجتمع Y

$$Y_{syw} = N\bar{y}_{syw} = N\sum_{i=1}^{n_s}\frac{\bar{y}_i}{n_s}$$

مع تقدير لتباين Y_{syw}

$$s^2(Y_{syw}) = N^2 s^2(\bar{y}_{syw})$$

و لمزيد من المعلومات مع بعض الأمثلة راجع Scheaffer, Retal (1996).

تمارين

1. لنفترض أن أحد البنوك لديه كشف بأسماء الزبائن الذين حصلوا على قرض من البنك، مسلسل بحسب تاريخ منح القرض من قبل البنك للعشرين سنة الماضية. هناك اتجاه لدى الزبائن بعدم دفع ما تبقى بذمتهم، وذلك لارتفاع تكاليف المعيشة للسنوات الأخيرة. يرغب البنك في تقدير مجموع المبالغ المستحقة وغير المدفوعة. ماذا تقترح له؟ هل يستخدم العينة العشوائية البسيطة أم العينة العشوائية المنتظمة ولماذا؟
2. تقوم إحدى الشركات بوضع أسماء العاملين فيها مرتبين حسب الأحرف الأبجدية ضمن كل فئة معينة لمقدار الدخل، وهذه الفئات تبدأ من أعلى الدخول إلى أقل الدخول، إذا كان الهدف تقدير معدل دخل العامل، هل نستخدم العينة العشوائية المنتظمة أم العينة العشوائية الطبقية، أم العينة العشوائية البسيطة لسحب عينة من المجتمع؟ وضح إجابتك.
3. قام قسم مراقبة الجودة في أحد المصانع باستخدام العينة المنتظمة، لتقدير معدل وزن المادة المستخدمة لصناعة علبة مشروبات غازية محمولة على حزام الإنتاج للمصنع. إن البيانات في الجدول الآتي تمثل عينة عشوائية منتظمة 1 in 50 لإنتاج المصنع ليوم واحد. قدر، ثم أوجد فترة 90% ثقة للوسط الحسابي . لنفرض أن N=1800.

وزن العلبة بالغرامات

11.80	12.01	12.03	12.01	11.97	12.00
11.83	12.00	11.98	12.03	11.98	11.91
11.88	11.90	11.87	11.98	12.01	11.87
11.89	11.94	11.93	11.91	11.87	12.05
12.05	11.93	11.97	11.95	11.93	11.72
12.04	12.02	12.05	11.87	11.98	11.85

4. استخدام البيانات في السؤال الثالث لتحديد حجم العينة المطلوب لتقدير الوسط الحسابي إذا كان حد الخطأ لتقدير الوسط الحسابي هو $d=0.7$ ودرجة ثقة 95%.

5. إذا كان قسم مراقبة الجودة في السؤال الثالث يهمه معرفة ما إذا كانت العلبة مطابقة للمواصفات الموضوعة لها أم لا ، فقام بفحص نفس العينة فوجد أن 6 علب في العينة لا تنطبق عليها المواصفات. قدر نسبة العلب الصالحة للاستعمال، ثم أوجد معامل التغير لتقدير النسبة.

6. يعطي الجدول الآتي أعداد العاملين في 11 شركة صناعية في إحدى المدن جرى سحبها باستخدام العينة المنتظمة من بين 300 شركة عاملة في المدينة.

الشركة	1	2	3	4	5	6	7	8	9	10	11
عدد العاملين	110	50	202	60	90	304	240	92	160	220	20

أ. قدر مجموع العاملين في القطاع الصناعي في هذه المدينة.

ب. أوجد تقدير التباين للمجموع الكلي ومن ثم أوجد 95% فترة ثقة للقيمة الحقيقية لعدد العاملين في القطاع الصناعي في هذه المدينة.

ت. قارن بين تقدير التباين في الفرع ب أعلاه وتقدير التباين للمجموع الكلي باستخدام العينة العشوائية البسيطة.

7. ترغب إحدى كليات المجتمع في تطوير علاقتها مع أهل المنطقة المحيطة بالكلية. قامت الكلية بسحب عينة منتظمة 1 in 300 من كشف بأسماء طلبة الكلية يحتوي على 4500 طالب وسألتهم عن مقدار ما صرفوا على الملابس في الفصل الماضي. البيانات الآتية تمثل مقدار ما صرفه الطلاب الذين تم سحبهم بالعينة:

30, 60, 20,1 5, 28,1 5, 24, 34,1 0, 50, 20, 25, 33, 48, 23

أ. قدر معدل مصروف طلبة الكلية على الملابس وأوجد فترة 95% ثقة للقيمة الحقيقية لمعدل المصروفات الشهرية على الملابس.

ب. قدر مجموع المصروف على الملابس من قبل طلبة الكلية، ثم أوجد فترة 95% ثقة للقيمة الحقيقية لمجموع المصروف على الملابس من قبل طلبة الكلية.

8. استخدم البيانات في السؤال السابق لتحديد حجم العينة لتقدير مجموع المصروف خلال الفصل ليكون ضمن 1000± وبدرجة ثقة 95%.

9. تتوقع دائرة السير أن تتجاوز إحدى نقاط التفتيش على الأقل N=4000 سيارة. حدد حجم العينة لتقدير النسبة P إذا كان حد الخطأ في التقدير d= 0.015 ودرجة الثقة 95%.

المراجع العربية

1. سليمان محمد طشطوش (2001)، أساسيات المعاينة الإحصائية، دار الشروق للنشر والتوزيع، عمان.
2. حسين علوان (1994)، طرق المعاينة، دار الفرقان، عمان.
3. هلال عبود البياتي وصبري رديف العاني (1981)، العينات، جامعة بغداد، بغداد.
4. وليد عبد الحميد نوري وعبد المجيد حمزة الناصر (1981) العينات، وزارة التعليم العالي والبحث العلمي، بغداد.
5. وليم كوكوران –ترجمة أنيس كنجو– تقنية المعاينة الإحصائية (1995)، جامعة الملك سعود، الرياض.

References

1. Barnett, V. (1991). Sampling Survey Principles and Methods, Edward Arnold, London.
2. Benedetto, J. J. and Ferreira, P. J. (2001). Modern Sampling Theory, Birkhauser Boston.
3. Bellhouse, L. D. (1988). Systematic Sampling. In Handbook of Statistics, Vol. 6, Sampling (Krishnaiah, P. R. and Rao, C. R., Eds.), 125–145, North Holland, Amsterdam.
4. Bellhouse, L. D. and Rao, J. N. K. (1975). Systematic Sampling in the Presence of a Trend. *Biometrika* 62, 694–697.
5. Brewer, K. W. R. (1963). A Model of Systematic sampling with Unequal Probabilities. *Australian Jour. Stat.* 5, 5–13.
6. Chaudhuri, A. and Stenger, H. (1992). Survey Sampling Theory and Methods, Marcel Dekker, New York.
7. Cochran, W. G. (1963). Relative Accuracy of Systematic and Stratified Random Samples for a Certain Class of Populations. *Ann. Math. Statist.* 17, 164–177.
8. Cochran, W. G. (1977). Sampling Technique, 3rd Ed. Wiley, New York.
9. Das, A. C. (1950). Two-dimensional Systematic Sampling and the Associated Stratified Random Sampling, *Sankhya*, 10, 95–108.
10. Deming, W. E. (1960). Sampling Design in Business Research. Wiley, New York.
11. Finney, D. J. (1948). Random and Systematic Sampling in Timber Surveys. *Forestry*, 22, 1–36.
12. Foreman, E. K. (1991). Survey Sampling Principles, Marcel Dekker, New York.
13. Gautschi, W. (1957). Some Remarks on Systematic Sampling. *Ann. Math. Statist.*, 28, 385–394.
14. Govindarajulu, Z. (1999). Elements of Sampling Theory and Methods, Prentice Hall.
15. Goodman, R. and Kish, L. (1950). Controlled Selection–a Technique in Probability Sampling. *J. Amer. Statist. Assoc.*, 45, 350–372.
16. Hajek, J (1981). Sampling from Finite Population, Marcel Dekker, New York.

17. Hansen, M. H., and Hurwitz, W. N. (1943). On the Theory of Sampling from Finite Populations. *Ann. Math. Statist.* 14, 333–362.
18. Iachan, R. (1982). Systematic Sampling a Critical Review. *Internat. Statist. Rev.* 50, 293–303.
19. Hedayat, A. S. and Sinha, B. K. (1991). Design and Inference in Finite Population Sampling, Wiley, New York.
20. Kish, L (1995). Survey Sampling, Wiley, New York.
21. Koop, J. C. (1963). On Splitting a Systematic Sample for Variance Estimation. *Ann. Math. Statist.* 42, 1084–1087.
22. Koop, J. C. (1988). The Technique of Replicated or Interpenetrating Samples, In: Handbook of Statistics, Vol. 6, Sampling (Krishnaiah, P. R. and Rao, C. R., Eds.), 125–145, North Holland, Amsterdam.
23. Kunte, S. (1978). A Note on Circular Systematic Sampling Desig, *Sankhya*, C40, 72–73.
24. Levy, P. S. and Lemeshow S. (1991). Sampling of Population: Methods and Applications, Wiley, New York.
25. Lohr, S. L. (1999). Sampling: Design and Analysis, Duxbury, New York.
26. Madow, W. G. and Madow, L. H. (1944). On the Theory of Systematic Sampling. *Ann. Math. Statist.* 15, 1–24.
27. Madow, L. H. (1946). Systematic Sampling and its Relation to other Sampling Designs. *J. Amer. Statist. Assoc.*, 41, 207–214.
28. Madow, L. H. (1949). In the Theory of Systematic Sampling II. *Ann. Math. Statist.* 20, 333–354.
29. Madow, W. G. (1953). On the Theory of Systematic Sampling III: Comparison of Centered and Random Start Systematic Sampling. *Ann. Math. Statist.* 24, 101–106.
30. Mahalanobis, P. C. (1946). Recent Experiments in Statistical Sampling in the Indian Statistical Institute. *J. Roy. Statist. Soc.* A109, 325–378, reprinted in Sankhyii.
31. Milne, A. (1959). The Centric Systematic Area Sample Treated as A Simple Random Sample. *Biometrics*, 15, 270–297.
32. Murthy, M. N. and Rao, T. J. (1988). Systematic Sampling with Illustrative Examples, In: Handbook of Statistics, Vol. 6, Sampling (Krishnaiah, P. R. and Rao, C. R., Eds.), 125–145, North Holland, Amsterdam.
33. Raj, D. (1968). Sampling Theory, McGraw Hill, New York.
34. Rao, J. N. K. (1985). Conditional Inference in Survey Sampling. Survey Methodology, D. I. *Inst. Statist. Math.* No.1, 15–31.
35. Sampath, S. (2000). Sampling Theory and Methods, CRC Press.
36. Scheaffer, R.L., Mendenhall, W. and Ott L. (1996). Elementary Sampling Survey, 5th ed., Duxbury, New York.
37. Singh, D., Jindal, K. and Grag, J. N. (1968).On Modified Systematic Sampling, *Biometrika*, 55, 541–546.
38. Singh, D. and Singh, P. (1977). New Systematic Sampling. *J. Statist. Plann. Infer.* 1, 163–178.
39. Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., and Ashok, C. (1984). Sampling Theory of Surveys with Applications, 3rd ed., Ames (Iowa): Iowa State University Press.
40. Thompson, S. K (2002). Sampling, 2end Wiley, New York.

41. Tornqvist, L. (1963). The Theory of Replicated Systematic Cluster Sampling with Random Start. *Rev. Internat. Statist. Inst.*, 31, 11–23.
42. Wu, C. F. (1984). Estimation in Systematic Sampling with Supplementary Observations. Sankhya, B46, 306–315.
43. Yates, F. (1981). Sampling Methods for Censuses and Surveys 4th Ed., Grittin, London.
44. Yates, F. (1948). Systematic Sampling, *Phil. Trans. Roy. Soc. London*, A 241, 345–371.
45. Zinger, A. (1964). Systematic Sampling in Forestry. *Biometrics*, 20, 553–565.

الفصل التاسع

العينة العنقودية بمرحلة واحدة
Single – Stage Cluster Sample

1.9 مقدمة

إن الهدف للمسح بالعينة مصمم للحصول على معلومات محددة عن معالم المجتمع بأقل كلفة ممكنة، وكما رأينا أن العينة العشوائية الطبقية أفضل من العينة العشوائية البسيطة للأسباب التي ذكرناها في حينها، سوف نتناول في هذا الفصل العينة العنقودية في مرحلة واحدة والتي في بعض الأحيان تعطي معلومات أكثر دقة لكل وحدة تكاليف من طرق المعاينة التي تكلمنا عنها سابقاً.

يمكن أن نعرف العينة العنقودية بمرحلة واحدة على أنها عينة عشوائية بسيطة، تكون فيها وحدات المعاينة عبارة عن مجموعة من العناصر أو عنقود منها، لقد رأينا أن العينة العشوائية البسيطة والعينة العشوائية الطبقية تحتاج إلى إطار المعاينة، ولكن في الغالب تكون عملية الحصول أو تكوين إطار للمعاينة شاقة ومكلفة كثيراً. وأحياناً لا يوجد إطار ولا يمكن تكوينه. لذا فإن السبب الأول لاستخدام العينة العنقودية هو عدم توافر إطار للمعاينة، أو لكون الحصول عليه يكلف كثيراً. أما السبب الثاني فهو أن تكاليف الحصول على المعلومات تزداد كلما تباعدت عناصر المجتمع عن بعضها الآخر.

توجد أمثله كثيرة لاستخدامات العينة العنقودية، سوف أكتفي بذكر مثال واحد سأشرح من خلاله عملية سحب العينة العنقودية في مرحلة واحدة. لنفرض أننا نريد أن نقدر معدل دخل العائلة في إحدى المدن، إن عملية إعداد إطار لجميع وحدات المجتمع التي تمثل هنا العائلات الساكنة في المدينة قد تكون صعبة أو مكلفة خصوصاً إذا كان عدد سكان المدينة كبيراً، غالباً نستطيع الحصول على كشف بقطاعات المدينة السكنية. لذا سوف نَعُدَّ أن كل قطاع يمثل وحدة من وحدات المجتمع، وطبعاً كل قطاع يحتوي على عدد من العائلات. يمكن أن نعرف عدد القطاعات الموجودة في المدينة، ولتقدير معدل دخل العائلة سوف نقوم بسحب عينة عشوائية بسيطة من القطاعات، ومن ثم نقوم بجمع معلومات عن جميع العائلات الموجودة في هذه القطاعات التي سحبت في العينة. ونستخدم هذه المعلومات لتقدير معدل دخل العائلة في المدينة. مما لا شك فيه أن العائلات الموجودة في كل قطاع ستكون متجاورة ومتجانسة تقريباً مع بعضها؛ لذا سيؤدي هذا إلى خفض تكاليف المعاينة.

29 تقدير الوسط الحسابي وا لمجموع الكلي للمجتمع

إن العينة العنقودية ما هي إلا عينة عشوائية بسيطة في كل وحدة من وحدات المعاينة توجد مجموعة من العناصر، لذلك سوف تكون عملية تقدير مَعَالم المجتمع مشابهة إلى حد ما لتلك التي حصلنا عليها في العينة العشوائية البسيطة، وخصوصاً تقدير الوسط الحسابي للمجتمع.

سوف نقوم باستخدام الرموز والمصطلحات الآتية في هذا الفصل

N : عدد العناقيد أو المجموعات في المجتمع.

n: عدد العناقيد أو المجموعات التي اختيرت باستخدام العينة العشوائية البسيطة.

m_i : عدد العناصر في العنقود i و $i=1, 2, \ldots, N$

$M=\sum_{i=1}^{N} m_i$: عدد العناصر في المجتمع.

$\overline{M}=M/N$: معدل حجم العنقود في المجتمع.

$\overline{m}=\frac{1}{n}\sum_{i=1}^{n} m_i$: معدل حجم العنقود في العينة.

y_i : مجموع المشاهدات للمتغير y في العنقود i.

1.29 تقدير الوسط الحسابي للمجتمع

سوف نستخدم الوسط الحسابي للعينة للمتغير y، والذي سنرمز له $\bar{y}_{cs}$ كتقدير للوسط الحسابي للمجتمع ويعرف

$$\bar{y}_{cs}=\frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} m_i}$$

أما تقدير تباين $\bar{y}_{cs}$ فيكون

$$s^2(\bar{y}_{cs})=\frac{N-n}{N\,n\,\overline{M}^2}\,\frac{\sum_{i=1}^{n}(y_i-m_i\,\bar{y}_{cs})}{n-1}$$

نلاحظ أن $\bar{y}_{cs}$ يأخذ شكل تقدير النسبة R والتي سبق أن تكلمنا عنها، كلُّ ما في الأمر استبدلنا m_i بدلاً عن x_i. لذلك من السهولة ان نتوصل إلى أن $\bar{y}_{cs}$ تقدير متحيز إلى $\overline{Y}$ الوسط الحسابي للمجتمع. وتباينه $var(\bar{y}_{cs})$ بأخذ شكل تباين النسبة R. إن تقدير التباين $s^2(\bar{y}_{cs})$ يعتبر تقديراً متحيزاً. لكن تقل كمية التحيز كلما كبر حجم العينة. هنا إذا كانت $n>20$ تعتبر

كبيرة. ولكن إذا كان حجم العناقيد متساوياً أي $m_1 = m_2 = ... = m_n = m$ يصبح تقديراً غير متحيز.

مثال (1): لتقدير المدخولات السنوية للعائلة في إحدى المدن، سحبت عينة عشوائية مكونة من 25 قطاعاً من بين 415 قطاعاً موجوداً في المدينة. استخدم البيانات الموجودة في الجدول الآتي لتقدير معدل دخل العائلة، وكذلك أوجد 95% فترة ثقة لمعدل لمدخولات العائلة السنوية.

مجموع الدخول (y_i)	عدد العائلات (m_i)	العنقود (i)	مجموع الدخول (y_i)	عدد العائلات (m_i)	العنقود (i)
49,000	10	14	96,000	8	1
53,000	9	15	121,00	12	2
50,000	3	16	42,000	4	3
32,000	6	17	65,000	5	4
22,000	5	18	52,000	6	5
45,000	5	19	40,000	6	6
37,000	4	20	75,000	7	7
51,000	6	21	65,000	5	8
30,000	8	22	45,000	8	9
39,000	7	23	50,000	3	10
47,000	3	24	85,000	2	11
41,000	8	25	43,000	6	13
			54,000	5	13

الحل:

$$\bar{y}_{cs} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} m_i} = \frac{1,329,000}{151} = 8,801$$

كذلك

$$\sum_{i=1}^{n}(y_i - m_i\bar{y}cs)^2 = \sum_{i=1}^{25} y_i^2 - 2\bar{y}_{cs}\sum_{i=1}^{25} y_i m_i + \bar{y}_{cs}^2\sum_{i=1}^{25} m_i^2$$

$$= 82,039,000,000 - 2(8,801)(8,403,000) + (8,801)^2(1,047)$$

$$= 15,227,502,247$$

بما أن $\bar{M}$ غير معروفة نقدرها باستخدام

$$\bar{m} = \frac{1}{n}\sum_{i=1}^{25} m_i = \frac{151}{25} = 6.04$$

$$s^2(\bar{y}_{cs}) = \frac{N-n}{N n \bar{M}^2}\frac{\sum_{i=1}^{n}(y_i - m_i\bar{y}_{cs})}{n-1} = \frac{(415-25)}{(415)(25)(6.04)^2}\frac{15,227,502,247}{24} = 653,785$$

لذا فإن تقدير الخطأ المعياري

$$s(\bar{y}_{cs}) = \sqrt{653,785} = 808.57$$

وكذلك فترة 95% ثقة تقريبية للوسط الحسابي $\bar{Y}$

$$\bar{y}_{cs} \pm z_{\alpha/2} s(\bar{y}_{cs})$$

$$8,801 \pm 1.96(808.57) \Rightarrow (7216.2, 10385.79)$$

نلاحظ أن فترة الثقة كبيرة.بالإمكان تصغيرها ؛ وذلك بزيادة حجم العينة.

2.2.9 تقدير المجموع الكلي للمجتمع

سوف نستخدم طريقتين لتقدير المجموع الكلي للمجتمع

1. **تقدير المجموع الكلي بالاعتماد على M**

سنرمز إلى تقدير المجموع الكلي هنا Y_{csm} ويعرف

$$Y_{csm} = M\bar{y}_{cs}$$

أما تقدير تباينه فيمكن إيجاده كما يأتي

$$s^2(Y_{csm}) = M^2 s^2(\bar{y}_{cs})$$

ملاحظة: إن هذا التقدير مفيدٌ إذا كان مجموع عناصر المجتمع M معلوم.

مثال (2): استخدم البيانات في مثال (1) لإيجاد تقدير المجموع الكلي للمجتمع. إذا علم أن هناك 2.500 عائلة في المدينة. وكذلك أوجد 90% فتره ثقة للمجموع الكليY.

$$Y_{csm} = M\bar{y}_{cs} = 2500(80801) = 22002,500$$

كذلك فإن 90% فترة ثقة للمجموع Y

$$(2,2002,500 \mp 3,325,243) \square$$
$$22,002,500 \pm 1.645\sqrt{(2,500)^2(653,785)}$$

2- **تقدير المجموع الكلي دون الاعتماد على M**

غالباً ما تكون عدد عناصر المجتمع غير معلومة، لذلك لا نستطيع استخدام Y_{csm} ، ولكن نستطيع إيجاد تقديرٍ آخر لا يعتمد علىM. لنفرض أن $\bar{y}_c$ على أنه معدل مجموع العناقيد في العينة ويعرف

$$\bar{y}_c = \frac{1}{n}\sum_{i=1}^{n} y_i$$

يعتبر $\bar{y}_c$ تقديراً غير متحيز إلى معدل مجموع العناقيد في المجتمع. كذلك يمكن أن نقدر المجموع الكلي Y بما يأتي

$$Y_{cs} = N\bar{y}_c = \frac{N}{n}\sum_{i=1}^{N} y_i$$

نلاحظ أن Y_{cs} تقدير غير متحيز إلى Y مع تقدير للتباين:

إذا

$$s^2(Y_{cs}) = N^2 s^2(\bar{y}_{cs}) = N^2\,\frac{N \square n}{Nn}\,\frac{\sum_{i=1}^{n}(y_i \square \bar{y}_c)^2}{n \square 1}$$

كان هناك تغيرٌ كبيرٌ بين حجم العناقيد ، وكان حجم العناقيد له علاقة قوية مع مجموع العناقيد سيكون تباين Y_{cs} بصورة عامة أكبر من تباين Y_{csm} . كذلك نلاحظ أن Y_{cs} لم يستخدم المعلومات المتوافرة حول حجم العناقيد $m_1, m_2, \ldots, m_n$ لذلك يعتبر أقل دقة من Y_{csm} الذي استخدم هذه المعلومات.

مثال (3): استخدم البيانات في مثال (1) لتقدير المجموع الكلي وكذلك أوجد فترة 90% ثقة للمجموع الكلي

الحل:

$$Y_{cs} = \frac{N}{n}\sum_{i=1}^{n} y_i = \frac{415}{25}(1,328,000) = 22,061,400$$

$$\sum_{i=1}^{N}(y_i - \bar{y}_c)^2 = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}(\sum_{i=1}^{n} y_i)^2 = 82,039,000,000 - \frac{1}{25}(1,329,000)^2$$
$$= 11,389,360,000$$

كذلك فإن فترة 90% ثقة للمجموع Y

$$Y_{cs} \pm z_{\alpha/2}\sqrt{s^2(Y_{cs})}$$

$$22,061,400 \pm 1.645\sqrt{(415)^2 \ \frac{415-25}{415(25)} \ \frac{11,389,360,000}{24}}$$

$$22,061,400 \pm 2,883,619$$

3.9 تقدير الوسط الحسابي والمجموع إذا كان حجم العناقيد متساوياً

توجد حالات كثيرة يكون فيها حجم العناقيد متساوياً أي $m_1 = m_2 = \ldots = m_n = m$ على سبيل المثال: يقوم أحد المصانع للمعلبات بوضع كل 64 علبة من إنتاجه في صندوق واحد ، فإذا كان إنتاج المصنع يومياً 1000

صندوق، وأرادت إدارة مراقبة الإنتاج أن تسحب عينة عشوائية مقدارها 30 صندوقاً يومياً وفحصها. نلاحظ هنا أن $N=1000$ و $n=30$ و $m_1 = m_2 = \ldots = m_{30} = 64$ و $M=mN=64000$.

سيكون تقدير الوسط الحسابي للمجتمع مع تقدير للتباين في هذه الحالة كما يأتي

$$\bar{y}_{cs} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} m_i} = \frac{1}{nm}\sum_{i=1}^{n} y_i$$

$$s^2(\bar{y}_{cs}) = \frac{N-n}{nNm^2} \frac{\sum_{i=1}^{n}(y_i - m\bar{y}_{cs})^2}{n-1}$$

أما تقدير المجموع الكلي ففي هذه الحالة $M=mN$ ستكون معلومة في الغالب. لذلك فإن تقدير المجموع الكلي سيكون

$$Y_{cs} = M\bar{y}_{cs} = M\frac{1}{nm}\sum_{i=1}^{n} y_i$$

أما تقدير التباين فسيكون

$$s^2(Y_{cs}) = M^2 s^2(\bar{y}_{cs})$$

مثال (4): أراد بائع صحف في إحدى المناطق السكنية المؤلفة من 4000 منزل أن يقدر عدد الصحف التي تشترى من قبل سكان كل منزل. إذا علمنا أن المنطقة مؤلفة من 400 قطاع، كل قطاع يحتوي على 10 منازل. قام بسحب عينة عشوائية مؤلفة من 4 قطاعات وحصل على النتائج الآتية. أوجد فترة 95% ثقة للوسط الحسابي $\bar{Y}$.

القطاع	عدد الصحف المشتراة من قبل المنازل المختلفة										المجموع
1	1	4	1	2	3	3	1	2	1	1	19
2	2	1	1	4	1	3	2	2	3	1	20
3	1	3	1	2	3	1	1	1	1	2	16
4	1	3	2	1	5	1	2	3	1	1	20

الحل:

$$\bar{y}_{cs}=\frac{1}{nm}\sum_{i=1}^{n}y_i\,\frac{19+20+16+20}{4(10)}=1.875$$

$$\sum_{i=1}^{4}(y_i-m\bar{y}_{cs})^2=\sum_{i=1}^{4}y_i^2-nm^2\,\bar{y}_{cs}^2$$

$$=(19)^2+(20)^2+(16)^2+(20)^2-4(10)^2(1.875)^2=10.75$$

$$s^2(\bar{y}_{cs})=\frac{N-n}{nNm^2}\;\frac{\sum_{i=1}^{n}(y_i-m\bar{y}_{cs})^2}{n-1}=\frac{(400-4)}{400(4)(10)^2}\;\frac{10.75}{3}=0.0089$$

لذا فإن فترة 95% ثقة للوسط الحسابي $\bar{Y}$

$$\bar{y}_{cs}\pm z_{\alpha/2}\sqrt{s^2(\bar{y}_{cs})}$$

$$1.875\mp 1.96\sqrt{0.0089}=\quad(1.875\mp 0.1849)$$

4.9 اختيار حجم العينة لتقدير الوسط الحسابي والمجموع الكلي

تتأثر كمية المعلومات في العينة العنقودية بعدد العناقيد وحجم العنقود النسبي. وكما لاحظنا فإن تقدير الخطأ المعياري يعتمد على مقدار التشتت بين مجاميع العناقيد، لذلك فإن أي محاوله لتقليل الخطأ المعياري لا بد فيها من اختيار عناقيد يكون التشتت بين مجاميعها قليلاً.

1.49 اختيار حجم العينة لتقدير الوسط الحسابي

سوف نفترض أن حجم العنقود (وحدات المعاينة) تم اختيارها وبقي أن نختار عدد العناقيد n يكمن كتابة تباين المجتمع إلى $\bar{y}_{cs}$

$$Var(\bar{y}_{cs}) = \frac{N-n}{nN\bar{M}^2}\, S_c^2$$

بحيث إن S_c^2 تباين المجتمع إلى

$$s_c^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - m_i\,\bar{y}_{cs})^2$$

بما أن S_c^2 غير معلومة سيتم تقديرها باستخدام s_c^2. حيث إن S_c^2 و $\bar{M}$ غير معلومات لذا فإن اختيار حجم العينة، أي عدد العناقيد تعتريه صعوبات، وللتغلب على هذه الصعوبات فإننا نستخدم الأسلوب نفسه الذي اتبعناه في تقدير النسبة R، وذلك بأن نقوم بتقدير S_c^2 و $\bar{M}$ إما باستخدام بيانات لتجارب سابقه مماثلة أو باستخدام عينة عشوائية أولية.

سوف نستخدم الأسلوب نفسه الذي سبق أن استخدمناه في اختيار حجم العينة حيث نقوم بحل المعادلة

$$2\sqrt{Var(\bar{y}_{cs})} = d$$

لإيجاد قيمة n. هنا d ترمز إلى حد خطأ التقدير. نستطيع إيجاد حجم العينة التقريبي لتقدير $\bar{Y}$ مع حد لخطأ التقدير مقداره d ودرجة ثقة 95% بما يأتي

$$n \geq \frac{NS_c^2}{ND + S_c^2}$$

حيث إننا نقدر S_c^2 و $\bar{M}$ باستخدام s_c^2 و $\bar{m}$ على التوالي، وكذلك فإن

$$D = \frac{d^2\,\bar{M}}{4}$$

مثال (5): استخدم الجدول في مثال (1) لإيجاد حجم العينة المطلوب بحيث يكون معدل دخل العائلة ضمن 500.

لا بد من تقدير S_c^2 باستخدام البيانات المتوافرة حول 25 عائلة باعتبارها عينة عشوائية أولية

$$s_c^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - m_i\,\bar{y}_{cs})^2 = 634,479,260$$

كذلك نقدر $\bar{M}$ باستخدام $\bar{m} = 6.04$. إذاً سيكون حجم D التقريبي

$$D = \frac{d^2\,\bar{m}^2}{4} = \frac{(500)^2(6.04)^2}{4} = 2,280,100$$

$$n \geq \frac{Ns_c^2}{ND + s_c^2} = \frac{(415)(634,479,260)}{(415)(2,280,100) + 634,479,260} = 167$$

2.4.9 اختيار حجم العينة لتقدير المجموع الكلي

كما مر معنا في حالة تقدير المجموع كان هناك طريقتان لتقدير المجموع الكلي. كذلك توجد طريقتان لاختيار حجم العينة لتقدير المجموع.

1. **عندما تكون M معلومة**

إذا كان حجم المجتمع M معلوماً فإن حجم العينة المطلوب لتقدير المجموع يكون

$$n \geq \frac{NS_c^2}{ND + S_c^2}$$

حيث إن S_c^2 تقدر باستخدام s_c^2 وذلك باستخدام معلومات سابقة أو سحب عينة عشوائية أولية وكذلك فإن

$$D = \frac{d^2}{4N^2}$$

مثال (6): استخدم الجدول في مثال (1) لإيجاد حجم العينة المطلوب لتقدير Y إذا كان M=2500 و d=1.000.000 ودرجة ثقة 95%.

الحل: أوجدنا في المثال السابق $s_c^2 = 634,479,260$ ، فإن

$$D = \frac{d^2}{4N^2} = \frac{(1,000,000)^2}{4(415)^2} = 1,415,589.5$$

$$n \geq \frac{Ns_c^2}{ND + s_c^2} = \frac{(415)(634,479,260)}{(1,451,589.5)(415) + 634,479,260} = 213$$

2. **عندما تكون M معلومة**

لقد سبق أن قدرنا تباين Y_{cs} باستخدام

$$s^2(Y_{cs}) = N^2 \frac{N-n}{nN} s_t^2$$

حيث إن

$$s_t^2 = \frac{1}{n-1} \sum_{i=1}^{n} (y_i - \bar{y}_c)^2$$

لذا فإن تباين المجتمع إلى Y_{cs} سيكون

$$Var(Y_{cs}) = N^2 \frac{N-n}{nN} S_t^2$$

حيث تم تقدير S_t^2 باستخدام s_t^2. إذاً نقدر Y ضمن حد خطأ للتقدير مقداره d يمكن إيجاد حجم العينة الذي يحقق لنا هذه الشروط بحل المعادلة

$$2\sqrt{Var(Y_{cs})} = d$$

لنحصل على

$$n \geq \frac{NS_t^2}{ND + S_t^2}$$

وبما أن S_t^2 غير معلومة نقدرها باستخدام s_t^2 وذلك باستخدام معلومات سابقة أو سحب عينة عشوائية أولية. كذلك فإن

$$D = \frac{d^2}{4N^2}$$

مثال (7):لنفرض أن M غير معلوم في مثال (1) كم حجم العينة الذي نحتاجه لتقدير المجموع إذا كان مقدار حد الخطأ المسموح به $d=1,000,000$

الحل: باستخدام المعلومات في الأمثلة السابقة نستطيع بسهولة إيجاد قيمة n كما يأتي

$$s_t^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y}_c)^2 = \frac{1}{24}(11,389,360,000) = 474,556,667$$

$$D = \frac{d^2}{4N^2} = \frac{(1,000,000)^2}{4(415)^2} = 2,903,179$$

$$n \geq \frac{NS_t^2}{ND + S_t^2} = \frac{(415)(474,556,667)}{(415)(2,903,179) + 474,556,667} = 183$$

5.9 تقدير النسبة

لنفرض أننا نرغب في تقدير نسبة المجتمع في مثالنا الخاص بتقدير معدل دخل العائلة ، يمكن أن يرغب الباحث في تقدير نسبة العائلات التي تعتمد في دخولها على فرد واحد من أعضاء العائلة. إن أفضل تقدير لنسبة المجتمع P هو نسبة العينة $\hat{p}_{(cs)}$ حيث تعرف

$$\hat{p}_{(cs)} = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} m_i}$$

a_i : ترمـز إلى مجمـوع العناصـر في العنقـود i ممّـن يحملـون صـفه معينـة و m_i يمثل عدد العناصر في العنقود.

ملاحظ: $\hat{p}_{(cs)}$ مشابه لـ $\bar{y}_{cs}$ الذي مر علينا سابقاً ما عدا أننا استخدمنا a_i بدلاً من y_i . إن تقدير تباين النسبة $\hat{p}_{(cs)}$ يكون

$$s^2(\hat{p}_{cs}) = \frac{N-n}{nN\bar{M}^2} \frac{\sum_{i=1}^{n}(a_i - m_i\,\hat{p}_{cs})^2}{n-1}$$

مثـال (8): عـودة إلى الجـدول الخـاص بالمثـال (1) إذا كانـت a_i تمثـل عـدد العـائلات في العنقـود i والـذين يعتمـدون في دخلـهم على فـرد واحـد مـن أعضـاء العائلة هي كما يأتي على التوالي:

1,7,4,3,0,4,3,3,1 ,3,2,4,1 ,4,5,2,3,1 ,2,3,2,4,4,3,3. أوجـد تقـدير النسـبة P، كذلك أوجد فترة 95% ثقة للنسبة لتقدير النسبةP.

الحل:

$$\hat{p}_{(cs)} = \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} m_i} = \frac{72}{151} = 0.477$$

لتقدير التباين لا بد من حساب

$$\sum_{i=1}^{n}(a_i - m_i\,\hat{p}_{cs})^2 = \sum_{i=1}^{n} a_i^2 - 2\hat{p}_{cs}\sum_{i=1}^{n} a_i m_i + \hat{p}_{cs}^2 \sum_{i=1}^{n} m_i^2$$

$$= 262 - 2(0.477)(511) + (0.477)^2(1047) = 12.729$$

نقدر $\bar{M}$ باستخدام

$$\bar{m} = \frac{1}{n}\sum_{i=1}^{n} m_i = \frac{151}{25} = 6.04$$

$$s^2(\hat{p}_{cs}) = \frac{415 - 25}{(415)(25)(6.04)^2} \frac{12.729}{24} = 0.00055$$

إن فترة 95% ثقة للنسبة لتقدير P

$$\hat{p}_{cs} \pm z_{\alpha/2}\sqrt{s^2(\hat{p}_{cs})}$$

$$0.477 \mp 1.96\sqrt{0.0055} \square \quad 0.477 \mp 0.046$$

6.9 اختيار حجم العينة لتقدير النسبة

إن إيجاد حجم العينة لتقدير النسبة P يمكن إيجاده باستخدام الطريقة نفسها التي أوجدنا فيها حجم العينة لتقدير $\bar{Y}$ ، لذا كان حد خطأ التقدير d ودرجة الثقة 95%، وذلك بحل المعادلة

$$2\sqrt{Var(\hat{p}_{cs})} = d$$

لإيجاد قيمة n

$$n \geq \frac{NS_c^2}{ND + S_c^2}$$

حيث إن

$$D = \frac{d^2\,\overline{M}^2}{4}$$

و S_c^2 يجري تقديرها باستخدام

$$s_c^2 = \frac{1}{n \square 1}\sum_{i=1}^{n}(a_i \square \hat{p}_{cs}\, m_i)^2$$

وذلك باستخدام معلومات سابقة أو سحب عينة عشوائية أولية.

مثال (9): عودة إلى المثال (8) أوجد حجم العينة المطلوبة لتقدير P ضمن حد لخطأ للتقدير مقداره 0.04 ودرجة ثقة 95%.

الحل: إن أفضل طريقة لتقدير S_c^2 باستخدام

$$s_c^2 = \frac{1}{n \square 1}\sum_{i=1}^{n}(a_i \square \hat{p}_{cs}\, m_i)^2 = \frac{12.729}{24} = 0.53$$

كذلك نقدر $\overline{M}$ باستخدام $\overline{m} = 6.04$ إذن

$$D = \frac{d^2 \overline{m}^2}{4} = \frac{(0.04)^2(6.04)^2}{4} = 0.0146$$

$$n \geq \frac{Ns_c^2}{ND + s_c^2} = \frac{(415)(0.53)}{(415)(0.0146) + 0.53} = 34$$

تمارين

1. ترغب شركة لإنتاج ماكينات الخياطة أن تقدر معدل تكاليف تصليح الماكينات التي تبيعها إلى مصانع مختلفة شهرياً، لا تستطيع أن تحصل على تكاليف التصليح لكل ماكينة، ولكن تستطيع الحصول على مجموع ما أنفقته على التصليح وعدد الماكينات الموجودة في كل مصنع يستخدم ماكيناتها. لذلك قررت أن عينة عنقودية تكون ملائمة، حيث إن كل مصنع يمثل عنقوداً. سحبت عينة عشوائية بسيطة حجمها 10 مصانع من 90 مصنعاً. يعطي الجدول الآتي عدد الماكينات ومجموع تكاليف التصليح لكل مصنع. قدر معدل تكاليف التصليح للماكينة وأوجد فترة 95% ثقة للوسط الحسابي للمجتمع $\overline{Y}$.

المصنع	1	2	3	4	5	6	7	8	9	10
عدد الماكينات	3	7	10	9	2	12	15	3	5	8
تكاليف التصليح	50	110	220	140	60	280	240	45	60	230

2. قدر مجموع ما تنفقه الشركة على التصليح ثم أوجد فترة 95% ثقة للمجموع الكلي لتكاليف التصليح في السؤال الأول.
3. بعد مراجعة سجلات البيع للشركة توصلوا إلى أن مجموع ما باعت الشركة للمصانع المختلفة في السؤال الأول هو 800 ماكينة خياطة. استخدم هذه المعلومات الجديدة لتقدير مجموع ما أنفقته الشركة على التصليح، ثم أوجد فترة 95% ثقة لمجموع تكاليف التصليح.
4. نفس الشركة في السؤال الأول ترغب بتحديد حجم العينة n لتقدير معدل تكاليف التصليح للشهر القادم. كم عدد العناقيد التي يجب أن

يختاروها لتقدير معدل تكاليف التصليح إذا كان حد الخطأ في التقدير $d=2$.

5. ترغب إدارة شؤون الطلاب تقدير نسبة الطلاب من أهل المدينة التي تتواجد فيها الجامعة والساكنين في سكن الجامعة. يتكون سكن الجامعة من بناية واحدة من أحد عشر طابقاً، كل طابق يحتوي على 70 غرفة. لنفرض أننا قمنا بسحب أربعة طوابق بطريقة عشوائية وحصلنا على النتائج الآتية:

الطابق	2	5	8	10
عدد الطلبة	10	12	5	17

قدر نسبة الطلاب من أهل المدينة والساكنين في سكن الجامعة ومن ثم أوجد فترة 95% ثقة للنسبة الحقيقية لعدد الطلاب الساكنين في سكن الجامعة من أهل المدينة التي توجد فيها الجامعة.

6. حدد أحد الباحثين الاقتصاديين بحثاً لتقدير معدل مصروفات العائلة على المنافع في إحدى المدن، ولكن لا يوجد كشف بجميع العائلات الساكنة في المدينة، لذلك قرر سحب عينة عشوائية بسيطة بحجم 11 قطاعاً من مجموع قطاعات المدينة البالغة 60 قطاعاً. قام العادون بجمع معلومات عن مصروفات العائلات على المنافع لكل القطاعات التي سحبت منها العينة. الجدول الآتي يعطينا عدد العائلات في كل قطاع ومجموع ما أنفق على المنافع. قدر معدل ما تنفقه العائلة على المنافع في المدينة، ثم أوجد معامل التغير لهذا التقدير.

القطاع	1	2	3	4	5	6	7	8	9	10	11
عدد العائلات	55	60	63	58	71	78	69	57	52	71	73
المجموع	2210	2390	2430	2380	2761	3000	2780	2370	2000	2800	2900

7. قدر مجموع ما تنفقه المدينة على المنافع، علماً بأن عدد العائلات القاطنة في المدينة غير معروف، ثم أوجد معامل التغير لهذا التقدير للبيانات في السؤال الخامس.
8. قرر الباحثون القيام بتقدير مجموع ما تنفقه مدينة مجاورة على المنافع مع حد خطأ في التقدير d=4000. استخدم البيانات الموجودة في السؤال الخامس لتحديد حجم العينة المطلوب للقيام بالبحث في المدينة المجاورة.
9. ترغب إحدى شركات سيارات الأجرة في إحدى المدن تقدير نسبة الإطارات الموجودة في سياراتها، والتي تُعد تالفة أو غير صالحة للاستعمال، إذا كان لدى الشركة 180 سيارة. قامت بسحب عينة عشوائية بسيطة بحجم n=25 باعتبار أن كل سيارة تمثل عنقوداً بأربعة إطارات. فوجدت عدد الإطارات غير الصالحة كما يأتي

2, 4, 0, 1, 0, 2, 4, 1, 3, 1, 2, 0, 1, 1, 2, 2, 4, 1, 0, 0, 3, 1, 2, 2, 1

قدر نسبة الإطارات غير الصالحة ثم أوجد فترة 95% ثقة للنسبة الحقيقية.

المراجع العربية

1. سليمان محمد طشطوش (2001)، أساسيات المعاينة الإحصائية، دار الشروق للنشر والتوزيع، عمان.
2. حسين علوان (1994)، طرق المعاينة، دار الفرقان، عمان.
3. هلال عبود البياتي وصبري رديف العاني (1981)، العينات، جامعة بغداد، بغداد.
4. وليد عبدالحميد نوري وعبدالمجيد حمزة الناصر (1981) العينات، وزارة التعليم العالي والبحث العلمي، بغداد.
5. وليم كوكوران –ترجمة أنيس كنجو– تقنية المعاينة الإحصائية (1995)، جامعة الملك سعود، الرياض.

References

1. Barnett, V. (1991). Sampling Survey Principles and Methods, Edward Arnold, London.
2. Chaudhuri, A. and Stenger, H. (1992). Survey Sampling□ Theory and Methods, Marcel Dekker, New York.
3. Cochran, W. G. (1977). Sampling Technique, 3rd Ed. Wiley, New York.
4. Cornfield, J. (1951). The Determination of Sample Size. *Amer. J. Publ. Health*, 41, 654□661.
5. Deming, W. E. (1960). Sampling Design in Business Research. Wiley, New York.
6. Durbin, J. and Sturat, A. (1954). Callbacks and Cluster Sampling Survey□An Experimental Study. *J. Roy. Stat. Soc.* A117, 387□428.
7. Foreman, E. K. (1991). Survey Sampling Principles, Marcel Dekker, New York.
8. Govindarajulu, Z. (1999). Elements of Sampling Theory and Methods, Prentice Hall.
9. Hajek, J (1981). Sampling from Finite Population, Marcel Dekker, New York.
10. Hansen, M. H., and Hurwitz, W. N. (1943). On the Theory of Sampling from Finite Populations. *Ann. Math. Statist.* 14, 333□362.
11. Hedayat, A. S. and Sinha, B. K. (1991). Design and Inference in Finite Population Sampling, Wiley, New York.
12. Kish, L (1957). Confidence Limits for Clustered Samples. *Amer. Soc. Rev.*, 22, 154□165.
13. Kish, L (1995). Survey Sampling, Wiley, New York.
14. Lahiri, D. B. (1951). A Method for Sample Selection Providing Unbiased Ratio Estimate. *Bull, Intern. Statist. Inst.* 31, 333□362.
15. Levy, P. S. and Lemeshow S. (1991). Sampling of Population□Methods and Applications, Wiley, New York.
16. Lohr, S. L. (1999). Sampling□Design and Analysis, Duxbury, New York.
17. Raj, D. (1968). Sampling Theory, McGraw Hill, New York.

18. Sampath, S. (2000). Sampling Theory and Methods, CRC Press.
19. Scheaffer, R.L., Mendenhall, W. and Ott, L. (1996). Elementary Sampling Survey, 5th ed., Duxbury, New York.
20. Sedransk, J. and Smith, P. J. (1988). Inference for Finite Population Quantiles. In Handbook of Statistics, Vol. 6, Sampling (Krishnaiah, P. R. and Rao, C. R., Eds.), 267–289, North Holland, Amsterdam.
21. Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., and Ashok, C. (1984). Sampling Theory of Surveys with Applications, 3rd ed., Ames (Iowa) Iowa State University Press.
22. Thompson, S. K. (2002). Sampling, 2end Wiley, New York.
Yates, F. (1981). Sampling Methods for Censuses and Surveys 4th Ed., Grittin, London.

الفصل العاشر

العينة العنقودية بمرحلتين
Two – Stage Cluster Sample

1.10 مقدمة

إن العينة العنقودية بمرحلتين ما هي إلا امتداد للعينة العنقودية بمرحلة واحدة، والتي تكلمنا عنها في الفصل السابق. وكما رأينا سابقاً فإن العنقود في بعض الأحيان عبارة عن تجمع طبيعي لمجموعة من العناصر، مثل قطاع يحتوي على مجموعة من المنازل أو صندوق يحتوي على مجموعة وحدات من إنتاج مصنع معين. يحتوي العنقود في أغلب الأحيان على مجموعة كبيرة من العناصر، يصعب جمع معلومات عن جميع هذه الوحدات، أو قد يحتوي العنقود على مجموعة من العناصر المتشابهة، لذلك فإن جمع معلومات عن مجموعة صغيرة من هذه العناصر يكون كافياً لإعطائنا معلومات عن العناصر داخل العنقود الواحد.

يمكن أن تُعرف العينة العنقودية بمرحلتين على أنها اختيار عينة عشوائية بسيطة من العناقيد، ومن ثم اختيار عينة عشوائية بسيطة من العناصر من جميع العناقيد التي اختيرت في المرحلة الأولى، فعلى سبيل المثال في مثالنا الذي تحدثنا عنه في الفصل السابق حول تقدير مصروفات العائلة في مدينة ما، حيث قسمنا المدينة إلى مجموعة من القطاعات (عناقيد) كل قطاع يحتوي على مجموعة من المنازل (العائلات)، سوف نقوم أولاً باختيار عينة

عشوائية بسيطة حجمه n من القطاعات (عناقيد)، ثم من كل قطاع تم اختياره نقوم بسحب عينة عشوائية بسيطة من المنازل، ونقوم بجمع معلومات من هذه المنازل.

هناك تشابه كبير بين العينة العنقودية بمرحلتين والعينة الطبقية. يمكن أن نعتبر المجتمع مقسماً إلى مجموعات من العناصر غير المتداخلة، إذا اعتبرنا هذه المجموعات طبقات، ومن ثم نختار عينة عشوائية بسيطة من كل مجموعة، ولكن إذا اعتبرنا هذه المجموعات عناقيد نختار عينة عشوائية بسيطة من بين هذه العناقيد، ومن ثم نختار عينة عشوائية بسيطة من بين عناصر كل عنقود جرى اختياره. إن العينة الطبقية تعطينا تقديرات إلى مَعَالم المجتمع بكون تباينها قليلاً، عندما يكون التغير بين عناصر المجموعة الواحدة قليلاً. أما العينة العنقودية بمرحلتين فتكون جيدة عندما يكون تغير العناصر داخل كل مجموعة كبيراً، والمجموعات (العناقيد) مشابه بعضها بعضًا.

أما مزايا العينة العنقودية بمرحلتين فهي ذاتها التي تكلمنا عنها في العينة العنقودية بمرحلة واحدة. ومن أهم هذه المزايا أنها لا تحتاج إلى إطار للمعاينة لتنفيذها، وكما نعرف فإن عملية إعداد إطار المعاينة مكلفة وشاقة كثيراً. وبالمقابل فإن عملية الحصول على كشف لجميع العناقيد الموجودة في المجتمع سهلة في الغالب. أما الميزة الأخرى فإن العينة العنقودية بمرحلتين تقلل تكاليف الحصول على المعلومات المطلوبة، وذلك لتقارب وحدات المعاينة في العينة العنقودية مما يقلل تكاليف السفر والوقت الذي يصرفه العادون في الحصول على المعلومات.

210 تقدير الوسط الحسابي والمجموع الكلي للمجتمع

سوف نستخدم الرموز والمصطلحات الآتية في هذا الفصل

N : عدد العناقيد في المجتمع.

n : عدد العناقيد في العينة التي اختيرت باستخدام العينة العشوائية البسيطة.

M_i : عدد العناصر في العنقود i.

m_i : عدد العناصر التي اختيرت من العنقود i باستخدام العينة العشوائية البسيطة.

$M = \sum_{i=1}^{N} M_i$: عدد العناصر في المجتمع.

$\overline{M}$: متوسط حجم العنقود في المجتمع.

y_{ij} : قيمة العنصر j المسحوب من العنقود i للمتغير y.

$\bar{y}_i = \frac{1}{m_i}\sum_{j=1}^{m_i} y_{ij}$: متوسط العينة للعنقود i.

1.210 تقدير الوسط الحسابي للمجتمع

لإيجاد تقدير الوسط الحسابي للمجتمع $\overline{Y}$ يمكن أن نستفيد من الطريقة التي استخدمناها في تقدير المجموع الكلي للمجتمع والتي لا تعتمد على M. نستطيع أن نحصل على تقدير غير متحيز للوسط الحسابي للمجتمع $\overline{Y}$ يمكن أن نعرفه

$$\bar{y}_{cs2} = \frac{N}{Mn}\sum_{i=1}^{n} M_i \bar{y}_i$$

أما تقدير تباين $\bar{y}_{cs2}$ فيكون

$$s^2(\bar{y}_{cs2}) = \frac{N-n}{N}\,\frac{s_b^2}{n\overline{M}^2} + \frac{1}{nN\overline{M}^2}\sum_{i=1}^{n} M_i^2\,\frac{M_i - m_i}{M_i}\,\frac{s_i^2}{m_i}$$

حيث إن

$$s_b^2 = \frac{1}{n-1}\sum_{i=1}^{n}(M_i\,\bar{y}_i - \bar{M}\,\bar{y}_{cs2})^2 ,\; s_i^2 = \frac{1}{m_i-1}\sum_{i=1}^{m_i}(y_{ij}-\bar{y})^2 ,\; i=1,2,\dots,n$$

نلاحظ أن $\bar{y}_{cs2}$ يعتمد على M عدد العناصر في المجتمع، وكذلك فإن s_i^2 عبارة عن تباين العينة العشوائية البسيطة المسحوبة من العنقود i.

مثال (1): توجد لدى إحدى شركات خياطة الملابس الجاهزة 20 مخيطة موزعة على أجزاء مختلفة من البلد. ترغب إدارة الشركة في تقدير معدل عدد الساعات التي تتوقف فيها ماكينات الخياطة عن العمل بسبب أعطال مختلفة شهرياً. بما أن المخيطات موزعة بصورة واسعة على البلد، وكل مخيطة تحتوي على عدد كبير من الماكينات، لذلك فإن عملية مشاهدة وفحص جميع سجلات الصيانة لكل الماكينات العاملة في الشركة تُعد عملية صعبة وتأخذ وقتاً طويلاً. لذلك قررت الشركة استخدام أسلوب المعاينة العنقودية بمرحلتين، في المرحلة الأولى جرى سحب 4 مخيطات بصورة عشوائية من بين العشرين مخيطة التابعة للشركة. أما المرحلة الثانية فقد جرى سحب عدد معين من الماكينات وبصورة عشوائية من كل مخيطة من المخيطات الأربع التي سحبت أولاً، ومن ثم جرت عملية مشاهدة سجلات الصيانة للماكينات التي سحبت من كل مخيطة، وحصلنا على النتائج المبينة في الجدول الآتي:

المخيطة	M_i	m_i	عدد ساعات التوقف شهرياً y_{ij}				$\bar{y}_i$	s_i^2
1	15	3	12	8	3		7.67	20.34
2	15	3	15	10	1		8.67	50.34
3	7	2	6	5			5.5	0.5
4	20	4	20	7	4	1	8	70

إذا كان عدد ماكينات الخياطة في جميع فروع الشركة $M=300$

الحل:

$$\bar{y}_{cs2} = \frac{N}{nM}\sum_{i=1}^{n} M_i \bar{y}_i = \frac{20}{4(400)}(443.5) = 7.39$$

$$s_b^2 = \frac{1}{n-1}\sum_{i=1}^{n}(M_i \bar{y}_i - \bar{M}\bar{y}_{cs2})^2 = 2678.84$$

$$\bar{M} = \frac{300}{20} = 15$$

$$s^2(\bar{y}_{cs2}) = \frac{N-n}{N}\ \frac{s_b^2}{n\bar{M}^2}\ \frac{1}{nN\bar{M}^2}\sum_{i=1}^{n} M_i^2\ \frac{M_i - m_i}{M_i}\ \frac{s_i^2}{m_i}$$

$$= \frac{20-4}{20}\ \frac{2678.84}{4(15)^2} + \frac{9849.55}{4(20)(15)^2} = 2.928$$

2210 تقدير المجموع الكلي للمجتمع

يمكن أن نحصل على تقدير غير متحيز إلى المجموع الكلي للمجتمع وذلك باستخدام $\bar{y}_{cs2}$ تقدير الوسط الحسابي للمجتمع، لذلك فإن تقدير المجموع الكلي وتقدير تباينه هما على التوالي

$$Y_{cs2} = M\bar{y}_{cs2} = \frac{N}{n}\sum_{i=1}^{n} M_i \bar{y}_i$$

$$s^2(Y_{cs2}) = \frac{N-n}{N}\ \frac{N^2}{n}s_b^2 + \frac{N}{n}\sum_{i=1}^{n} M_i^2\ \frac{M_i - m_i}{M_i}\ \frac{s_i^2}{m_i}$$

نلاحظ أننا لا نحتاج لمعرفة M عدد عناصر المجتمع للحصول على تقدير المجموع الكلي للمجتمع.

مثال (2): أوجد المجموع الكلي وتقدير الخطأ المعياري لتقدير المجموع الكلي للبيانات في المثال السابق.

الحل :

$$Y_{cs2} = M\bar{y}_{cs2} = 300(7.39) = 2217$$

$$s^2(Y_{cs2}) = M\sqrt{s^2(\bar{y}_{cs2})} = 300\sqrt{2.928} = 513.376$$

3.10 تقدير النسبة للوسط الحسابي

لقد لاحظنا أن تقدير الوسط الحسابي للمجتمع قبل قليل يعتمد على M عدد عناصر المجتمع، ولكن في كثير من الأحيان قد لا تكون عدد عناصر المجتمع معلومة، لذلك لا بد من تقدير الوسط الحسابي $\bar{Y}$ باستخدام المعلومات المتوافرة لدينا من العينة، وذلك عن طريق تقدير M، حيث يمكن أن نقدره باستخدام

$$M = N\sum_{i=1}^{n}\frac{M_i}{n}$$

إذا عوضنا عن قيمة M باستخدام M في $\bar{y}_{cs2}$ نحصل على ما يسمى بتقدير النسبة

$$\bar{y}_{csR} = \frac{\sum_{i=1}^{n} M_i \bar{y}_i}{\sum_{i=1}^{n} M_i}$$

كذلك فإن تقدير التباين إلى $\bar{y}_{csR}$

$$s^2(\bar{y}_{csR}) = \frac{N-n}{N}\,\frac{s_t^2}{n\,\bar{M}^2} + \frac{1}{nN\,\bar{M}^2}\sum_{i=1}^{n} M_i^2\,\frac{M_i - m_i}{M_i}\,\frac{s_i^2}{m_i}$$

حيث إن

$$s_t^2 = \frac{1}{n-1}\sum_{i=1}^{n} M_i^2(\bar{y}_i - \bar{y}_{csR})^2 \quad , \quad s_i^2 = \frac{1}{m_i - 1}\sum_{j=1}^{m_i}(y_{ij} - \bar{y}_i)^2$$

نلاحظ أن $\bar{y}_{csR}$ هو تقدير متحيز إلى $\bar{Y}$، كذلك فإن تقدير تباينه يعتمد على $\bar{M}$ حيث يمكن تقديره باستخدام $\bar{m}$. لا بد من الإشارة هنا إلى أن قيمة التحيز تصبح صغيرة وغير مهمة كلما كبر حجم العينة.

مثال (3): إذا كانت M غير معلومة في مثال (1) قدر معدل عدد ساعات التوقف لماكينات الخياطة في الشركة، كذلك أوجد معامل التغير.

الحل:

$$\bar{y}_{csR}=\frac{\sum_{i=1}^{n}M_i\bar{y}_i}{\sum_{i=1}^{n}M_i}=\frac{443.6}{57}=7.78$$

$$\bar{m}=\frac{1}{n}\sum_{i=1}^{n}M_i=\frac{57}{4}=14.25$$

$$s_t^2=\frac{1}{n-1}\sum_{i=1}^{n}M_i^2(\bar{y}_i-\bar{y}_{csR})^2=\frac{455.027}{3}=151.676$$

$$s^2(\bar{y}_{csR})=\frac{(20-4)}{20(4)(14.25)^2}(151.676)+\frac{9849.55}{4(20)(14.25)^2}=0.0755$$

$$CV(\bar{y}_{csR})=\frac{\sqrt{s^2(\bar{y}_{csR})}}{\bar{y}_{csR}}=\frac{\sqrt{0.0755}}{7.78}=0.035$$

410 تقدير نسبة المجتمع

إن تقدير النسبة في المجتمع يستخدم كثيراً في الحياة العملية، ففي مثالنا في بداية هذا الفصل حول تقدير مصروفات العائلات يمكننا تقدير نسبة العائلات التي تعتمد في مصروفاتها الشهرية على فرد واحد من أفراد العائلة. لذلك لا بد من إعادة تعريف y_{ij} على الوجه الآتي

$y_{ij}=1$ إذا كان العنصر j في العنقود i يحمل صفة مميزة

$y_{ij}=0$ خلاف ذلك أي أن العنصر لا يحمل الصفة المميزة

بما أنه في العادة M غير معلومة سوف نقدر النسبة P دون الاعتماد على قيمة M. إذا كان $\hat{p}_i$ يمثل نسبة الذين يحملون صفة معينة أو ينتمون إلى فئة معينة في العنقود i فإن تقدير النسبة P للمجتمع يمكن إيجاده باستخدام

$$\hat{p}_{csR}=\frac{\sum_{i=1}^{n}M_i\hat{p}_i}{\sum_{i=1}^{n}M_i}$$

وكذلك فإن تقدير تباين النسبة

$$s^2(\hat{p}_{csR}) = \frac{N-n}{N}\,\frac{s_t^2}{n\,\overline{M}^2} + \frac{1}{nN\,\overline{M}^2}\,\sum_{i=1}^{n} M_i^2\,\frac{M_i - m_i}{M_i}\,\frac{\hat{p}_i q_i}{m_i - 1}$$

حيث إن

$$s_t^2 = \frac{1}{n-1}\sum_{i=1}^{n} M_i^2(\hat{p}_i - \hat{p}_{csR})^2$$

مثال (4): عودة إلى مثال (1) إذا كانت نسبة الماكينات التي تحتاج إلى صيانة كبيرة، وهي على التوالي: $\hat{p}_1 = 0.4, \hat{p}_2 = 0.3, \hat{p}_3 = 0.25, \hat{p}_4 = 0.15$. حيث إن $\hat{p}_i$ و$i=1,2,3,4$ يمثل النسبة في العنقود i .أوجد $\hat{p}_{csR}$ وتقدير خطئه المعياري إذا كانت M غير معلومة.

الحل:

$$\hat{p}_{csR} = \frac{\sum_{i=1}^{n} M_i\,\hat{p}_i}{\sum_{i=1}^{n} M_i} = \frac{125.25}{57} = 0.267$$

$$\overline{m} = 14.25$$

$$s_t^2 = \frac{1}{n-1}\sum_{i=1}^{n} M_i^2(\hat{p}_i - \hat{p}_{csR})^2 = \frac{9.7846}{3} = 3.26$$

$$s^2(\hat{p}_{csR}) = \frac{20-4}{20}\,\frac{3.26}{4(14.25)^2} + \frac{60.66}{4(20)\,(14.25)^2} = 0.0069$$

$$s(\hat{p}_{csR}) = \sqrt{s^2(\hat{p}_{csR})} = \sqrt{0.0069} = 0.083$$

5.10 اختيار حجم العينة

إن اختيار حجم العينة في العينة العنقودية بمرحلتين يُعد أكثر تعقيداً منه في العينة العنقودية بمرحلة واحدة. إذ لا بد من اختيار عدد العناقيد n، وكذلك لا بد من اختيار حجم العينة من كل عنقود m_i و $i=1,2,\ldots,n$. إن اختيار n و m_i يعتمد على مصدرين من مصادر التغير الأول بين العناقيد والثاني داخل كل عنقود. بصورة عامة نسحب حجم عينة أكبر من العناقيد ذات التغير الكبير، إذا كانت العناقيد متجانسة داخلياً، أي عناصر كل عنقود متقاربة من بعضها بعضا، ولكن إذا كان هنالك اختلافات كبيرة بين أوساط العناقيد المختلفة، نقوم بسحب عناصر قليلة من كل عنقود، وبالمقابل نرفع عدد العناقيد المسحوبة (أي نزيد حجم n ونقلل حجم m_i). وإذا كانت هنالك اختلافات كبيرة داخل كل عنقود، ولكن أوساط العناقيد متقاربة، نقوم بزيادة العناصر المسحوبة من كل عنقود ونقلل عدد العناقيد (أي نزيد حجم m_i ونقلل حجم n).

لنفرض أن أحجام العناقيد متساوية أي $M_1 = M_2 = L = M_N = \overline{M}$ كذلك $m_1 = m_2 = \ldots = m_n = m$. إذا عوضنا عن قيمة M_i و m_i كذلك أهملنا معامل التصحيح فإننا سنحصل على تقدير إلى الوسط الحسابي $\overline{y}_{cs_2}$ للمجتمع وتباينه على التوالي

$$\overline{y}_{cs_2} = \frac{1}{n}\sum_{i=1}^{n}\overline{y}_i$$

$$Var(\overline{y}_{cs_2}) = \frac{S_b^2}{n} + \frac{S_w^2}{nm}$$

حيث إن S_b^2 التباين بين أوساط العناقيد و S_w^2 التباين بين العناصر داخل العناقيد. لإيجاد حجم العينة n و m إما أن نصغر التباين $Var(\overline{y}_{cs_2})$ عندما تكون التكاليف ثابتة، أو نقلل التكاليف عندما يكون التباين مساوياً إلى

كمية ثابتة. لا بد من معرفة التكاليف، لذلك نفرض أن تكاليف معاينة العنقود الواحد C_1 وتكاليف معاينة كل عنصر داخل العناقيد C_2 ومجموع التكاليف $C = nC_1 + nmC_2$. إن قيمة m التي تصغر $var(\bar{y}_{cs2})$ إذا كانت التكاليف C ثابتة أو نصغر C إذا كان التباين ثابتاً هي

$$m = \sqrt{\frac{C_1\, S_w^2}{C_2\, S_b^2}}$$

بعد أن نحدد قيمة m (عدد العناصر التي سوف نختارها من كل عنقود) نقوم بتحديد عدد العناقيد n إذا كان التباين ثابتاً باستخدام

$$var(\bar{y}_{cs2}) = \frac{S_b^2}{n} + \frac{S_w^2}{nm}$$

أو إذا كانت التكاليف ثابتة نستخدم

$$C = nC_1 + nmC_2$$

نلاحظ أن S_w^2 و S_b^2 غير معروفتين لذلك لا بد من تقديرهما إما باستخدام مسوحات مماثلة سابقة، أو بسحب عينة عشوائية أولية، واستخدام البيانات في العينة الأولية لإيجاد تقدير غير متحيز إلى S_w^2 هو

$$s_w^2 = \frac{1}{n}\sum_{i=1}^{n} s_i^2$$

ومن ثم نجد تقديرًا إلى S_b^2 هو

$$s_b^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\bar{y}_i - \bar{y}_{cs2})^2 - \frac{s_w^2}{m}$$

مثال (5): لمثالنا (1) لنفرض أن M=15 وm=4 كذلك $\bar{y}_1 = 13.4$ و $\bar{y}_2 = 8.67$ و $\bar{y}_3 = 4$ و $\bar{y}_4 = 10.34$ ، وكذلك فإن $s_1^2 = 20.34$ و $s_2^2 = 50.38$ و $s_3^2 = 7$ و $s_4^2 = 77.34$. أوجد قيمة n وm إذا ك أن $var(\bar{y}_{cs2}) = 2$ و $C_1 = 9$ و $C_2 = 1$ وC=250.

الحل :

$$s_w^2 = \frac{1}{4}\sum_{i=1}^{4} s_i^2 = 37.51 \, , \; \bar{y}_{cs2} = \frac{1}{4}\sum_{i=1}^{4} \bar{y}_i = 7.67$$

$$s_b^2 = \frac{1}{3}\sum_{i=1}^{4}(\bar{y}_i - \bar{y}_{cs2})^2 - \frac{s_w^2}{4} = 15.41 - \frac{37.51}{4} = 6.03$$

$$m = \sqrt{\frac{C_1 S_w^2}{C_2 S_b^2}} = \sqrt{\frac{(9)(7.67)}{(1)(6.03)}} = 4$$

إذا كان التباين إلى $\bar{y}_{cs2}$ ثابتاً ومساوياً إلى 2 أي $Var(\bar{y}_{cs2}) = 2$ نجد قيمة n باستخدام

$$Var(\bar{y}_{cs2}) = \frac{S_b^2}{n} + \frac{S_w^2}{4n} = 2 \quad or \quad 6.03 + \frac{37.51}{4} = 2n$$

إذاً $n = 8$

إذا كانت $C=250$ ، نجد قيمة n باستخدام

$$C = nC_1 + nmC_2 \quad or \quad 250 = n + 36n$$

إذاً $n = 7$

تمارين

1. يرغب صاحب مشتل في تقدير معدل أطوال الشتلات الموجودة، في حقل مقسم إلى 50 قطعة تختلف قليلاً في الحجم. لذلك قرر أن يسحب عينة بحجم 10% من كل قطعة باستخدام العينة العنقودية بمرحلتين، الجدول الآتي يعطينا البيانات اللازمة.

القطعة	عدد الشتلات	عدد الشتلات في العينة	أطوال الشتلات في العينة
1	52	5	12,11,12,10,13
2	56	6	10,9,7,9,10,8
3	60	6	6,5,7,5,4,6
4	46	5	7,8,7,7,8
5	49	5	10,11,15,12,11
6	51	5	14,15,13,12,13

قدر معدل أطوال الشتلات في الحقل ثم أوجد فترة 95% ثقة للوسط الحسابي للمجتمع.

2. في السؤال الأول لنفرض عدد الشتلات الموجودة في الحقل 2500، استخدم هذه المعلومات الإضافية لتقدير معدل أطوال الشتلات، ثم أوجد معامل التغير للتقدير.

3. لدى إحدى الشركات مجموعة كبيرة من المطاعم موزعة على 20 مدينة في إحدى البلدان، ترغب الشركة معرفة نسبة المطاعم التي لا تلتزم بشروط النظافة المطلوبة من قبل الشركة. المطاعم الموجودة في المدينة الواحدة يبدو أنها تُظهر نفس المستوى من النظافة. لذلك قررت إدارة الشركة سحب عينة عنقودية بمرحلتين، قامت بسحب 5 مدن بطريق عشوائية، ثم سحبت نصف المطاعم الموجودة في هذه المدن. يوضح الجدول الآتي البيانات المطلوبة. قدر نسبة المطاعم التي تمتلكها الشركة والتي لا

تلتزم بشروط النظافة المطلوبة من قبل الشركة، ثم أوجد فترة 90% ثقة للنسبة الحقيقية.

المدينة	عدد المطاعم في العينة	عدد المطاعم في المدينة	عدد المطاعم التي لا تلتزم بشروط النظافة
1	4	9	1
2	5	10	2
3	7	15	3
4	6	12	2
5	3	6	1

4. إذا علمنا أن عدد المطاعم التي تمتلكها الشركة في السؤال الثالث 400 مطعم. قدر نسبة المطاعم التي لا تلتزم بشروط النظافة، ثم أوجد معامل التغير لهذا التقدير.

5. يرغب أحد الباحثين الاجتماعيين في تقدير عدد المتقاعدين في إحدى المدن، لذلك قرر أن يستخدم العينة العنقودية بمرحلتين، وذلك بسحب عينة من قطاعات المدينة أولاً، ومن كل قطاع يسحب عينة من العائلات، علماً بأنه يوجد في المدينة 300 قطاع. قدر مجموع المتقاعدين القاطنين في هذه المدينة، ثم أوجد فترة 95% ثقة للمجموع الحقيقي.

قطاع	عدد العائلات	عدد العائلات في العينة	عدد المتقاعدين
1	18	3	1, 0, 2
2	4	3	0, 3, 0
3	9	3	1, 1, 2
4	12	3	0, 1, 1

6. استخدم البيانات في السؤال الخامس لتقدير معدل عدد المتقاعدين في العائلة، ثم أوجد معامل التغير لهذا التقدير.

7. يحتوي أحد المجتمعات على 10.000 عنصر. قسم المجتمع إلى 50 مجموعة جزئية حسب إحدى المواصفات. سحبت عينة عنقودية بمرحلتين: في المرحلة الأولى جرى سحب عينة من أربع مجموعات،

ومن كل مجموعة جرى سحب 40 عنصراً. الجدول الآتي يلخص النتائج التي حصلنا عليها.

المجموعة	N_i	n_i	$\bar{y}_i$	s_i^2	$\hat{p}_i$
1	150	50	142	540	0.42
2	250	50	154	482	0.38
3	200	50	130	623	0.36
4	180	50	165	508	0.32

احسب تقدير معدل المجتمع، وتقدير النسبة، ومن ثم أوجد فترة 95% ثقة لكلا التقديرين.

المراجع العربية

1. سليمان محمد طشطوش (2001)، أساسيات المعاينة الإحصائية، دار الشروق للنشر والتوزيع، عمان.
2. حسين علوان (1994)، طرق المعاينة، دار الفرقان، عمان.
3. هلال عبود البياتي وصبري رديف العاني (1981)، العينات، جامعة بغداد، بغداد.
4. وليد عبدالحميد نوري وعبدالمجيد حمزة الناصر (1981) العينات، وزارة التعليم العالي والبحث العلمي، بغداد.
5. وليم كوكوران –ترجمة أنيس كنجو– تقنية المعاينة الإحصائية (1995)، جامعة الملك سعود، الرياض.

References

1. Barnett, V. (1991). Sampling Survey Principles and Methods, Edward Arnold, London.
2. Brewer, K. W. R. and Hanif, M. (1970). Durbin's new Multistage Variance Estimator. *J. Roy. Stat. Soc.* B32, 302–311.
3. Chaudhuri, A. and Stenger, H. (1992). Survey Sampling: Theory and Methods, Marcel Dekker, New York.
4. Cochran, W. G. (1977). Sampling Technique, 3rd Ed. Wiley, New York.
5. Cornfield, J. (1951). The Determination of Sample Size. *Amer. J. Publ. Health*, 41, 654–661.
6. Deming, W. E. (1960). Sampling Design in Business Research. Wiley, New York
7. Foreman, E. K. (1991). Survey Sampling Principles, Marcel Dekker, New York.
8. Govindarajulu, Z. (1999). Elements of Sampling Theory and Methods, Prentice Hall.
9. Hajek, J (1981). Sampling from Finite Population, Marcel Dekker, New York.
10. Hansen, M. H., and Hurwitz, W. N. (1943). On the Theory of Sampling from Finite Populations. *Ann. Math. Statist.* 14, 333–362.
11. Hedayat, A. S. and Sinha, B. K. (1991). Design and Inference in Finite Population Sampling, Wiley, New York.
12. Kish, L and Hess, I. (1959). On Variance of Rations and their Differences in multistage samples, *J. Amer. Stat. Assoc.* 54, 416–446.
13. Kish, L (1995). Survey Sampling, Wiley, New York.
14. Lahiri, D. B. (1951). A Method for Sample Selection Providing Unbiased Ratio Estimate, *Bull, Intern. Statist. Inst.* 31, 333–362.
15. Levy, P. S. and Lemeshow S. (1991). Sampling of Population: Methods and Applications, Wiley, New York.
16. Lohr, S. L. (1999). Sampling: Design and Analysis, Duxbury, New York.

17. Murthy, M. N. (1967). Sampling Theory and Methods. Calcutta, India: Statistical Publication Society.
18. Raj, D. (1966). Some Remarks on a Simple Procedure for Sampling without Replacement, *J. Amer. Statist. Assoc.* 61, 391–397.
19. Raj, D. (1968). Sampling Theory, McGraw Hill, New York.
20. Sampath, S. (2000). Sampling Theory and Methods, CRC Press.
21. Scheaffer, R.L., Mendenhall, W. and Ott L. (1996). Elementary Sampling Survey, 5th ed., Duxbury, New York.
22. Sedransk, J. and Smith, P. J. (1988). Inference for Finite Population Quantiles. In: Handbook of Statistics, Vol. 6, Sampling (Krishnaiah, P. R. and Rao, C. R., Eds.), 267–289, North Holland, Amsterdam.
23. Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., and Ashok, C. (1984). Sampling Theory of Surveys with Applications, 3rd ed., Ames (Iowa): Iowa State University Press.
24. Thompson, S. K. (2002). Sampling, 2end Wiley, New York.
25. Yates, F. (1981). Sampling Methods for Censuses and Surveys 4th Ed., Grittin, London.

الفصل الحادي عشر

المعاينة مختلفة الاحتمالات
Varying Probability Sampling

1.11 مقدمة

لقد تكلمنا في الفصول العشر الأولى من هذا الكتاب عن اختيار وحدات العينة من المجتمع، بحيث إن جميع وحدات المجتمع لها الفرصة نفسها في الظهور في العينة، كما هي الحالة في العينة العشوائية البسيطة أو الطبقية أو المنتظمة ...إلخ. ولكن عندما تكون وحدات المعاينة مختلفة كثيراً في الحجم فإن العينة العشوائية البسيطة أو المنتظمة لا تعطينا تقديراً جيداً، وذلك بسبب التغير الكبير في الوحدات بالنسبة إلى الصفة y التي تجري مشاهدتها أو دراستها، فعلى سبيل المثال إذا كانت الوحدة تمثل حقولاً زراعية أو مدناً واضح أن مساحات الحقول لا تكون متساوية وكذلك عدد سكان المدن المختلفة لا يكون متساوياً. يمكن في مثل هذه الحالات أن نحصل على تقدير أفضل لو أعطينا احتمالاً أكبر لاختيار الوحدات الكبيرة. يمكن اختيار العينة مع الإرجاع أو دون إرجاع، هذا النوع من طرق المعاينة يسمى

Sampling with Probability proportional to size

أو(PPS Sampling) سوف نستخدم الرمز PPS في هذا الفصل للدلالة على أن وحدات العينة سحبت باحتمال يتناسب مع حجمها. كذلك سوف يقتصر كلامنا في هذا الفصل فقط على ما يسمى بالمعاينة مع احتمال يتناسب مع الحجم (PPS) سوف لا نتطرق إلى طرق أخرى يكون احتمال سحب الوحدات

من المجتمع غير متساوٍ. لتوضيح الفكرة؛ لنفرض أننا نرغب في تقدير عدد فرص العمل المتوافرة في الشركات العاملة في البلد، واضح أن الشركات ذات الحجم الكبير يكون توافر فرص العمل فيها أكبر من الشركات الصغيرة، ومن ثم لو أننا سحبنا عينة عشوائية بسيطة من الشركات دون التمييز بينها حسب حجمها فإن تقديرنا لعدد فرص العمل لا يكون دقيقاً، وذلك لأن التباين سوف يكون كبيراً، لأن التغير كبير بين حجم الوحدات والتي تمثل هنا شركات عاملة في البلد.

2.11 العينة العشوائية البسيطة مع الإرجاع

لنفرض أن لدينا مجتمعاً عدد وحداته N ،ولنفرض أننا نرغب في دراسة الصفة y للوحدة i و $i=1, 2, \ldots, N$ ولنفرض كذلك أنَّ احتمال سحب الوحدات من المجتمع يتناسب مع حجمها X_i أي $p_i = X_i/X$ ، بحيث إن $X = \sum_{i=1}^{N} X_i$.

وكذلك نرفض أننا سوف نختار وحدات العينة n مع الإرجاع، وكذلك لنفرض أن (y_i, p_i) يمثل القيمة والاحتمال للوحدة i و $i= 1, 2, \ldots, n$ من العينة. ويمكن ملاحظة أن y_i/p_i و $i=1, 2, \ldots, n$ مستقلة ولها نفس التوزيع. إذا كان $p_i = 1/N$ تكون العينة عشوائية بسيطة. ولهذا نلاحظ أن العينة العشوائية البسيطة ما هي إلا حالة خاصة من العينة المختلفة الاحتمالات.

يمكن الحصول على تقدير غير متحيز لتقدير المجموع الكلي Y للمجتمع وباحتمالات متناسبة مع حجم الوحدة (PPS) كما يأتي:

$$Y_{PPS} = \frac{1}{n} \sum_{i=1}^{n} y_i/p_i$$

مع تباين

$$Var(Y_{PPS}) = \frac{1}{n} \sum_{i=1}^{n} p_i (y_i/p_i - Y)^2$$

وتقدير تباين

$$s^2(Y_{PPS}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} p_i (y_i/p_i - Y_{PPS})^2$$

ويمكن مقارنة هذه المعاينة بالعينة العشوائية البسيطة لنحصل على $Var(Y_{PPS}) \leq Var(Y)$. لبرهنة هذه النتيجة ولمزيد من التفصيلات حول هذا النوع من المعاينة يراجع Singh and Chaudhary (1986).

3.11 العينة العشوائية البسيطة دون الإرجاع

لقد سبق وتكلمنا عن سحب عينة عشوائية بسيطة ودون إرجاع من المجتمع، وهذه الطريقة تعطي جميع وحدات المجتمع نفس الفرصة في الظهور في العينة. ولكن في بعض الحالات يكون إعطاء احتمالٍ متساوٍ لجميع وحدات المجتمع بالظهور في العينة غير ملائم، وذلك لكون وحدات المجتمع غير متساوية أو مختلفة في الحجم، والحجم له تأثير مباشر على قيمة المشاهدة في الوحدة.

سوف نحاول توضيح فكرة سحب الوحدة من المجتمع مع احتمال تناسب مع حجم الوحدة. لنفرض أننا نرغب في تقدير مجموع محصول القمح للحقول المزروعة بالقمح في إحدى المحافظات، وقمنا بإعداد كشف بجميع الحقول المزروعة بالقمح في هذا الفصل في المحافظة، نستطيع أن نسحب عينة عشوائية بسيطة. وباحتمال سحب متساوٍ لجميع الحقول كما مر معنا في الفصل الخامس. ولكن واضح أن الوحدة هنا تمثل حقلاً والحقول غير متساوية في مقدار المساحة المزروعة، ومن ثم فإن إعطاء احتمال متساوٍ لجميع الحقول بالظهور في العينة لا يكون ملائماً في مثل هذه الحالة، لكون مساحة الحقل لها علاقة مباشرة بمقدار القمح الذي سنحصل عليه من الحقل. لنفرض أن

لدينا خارطة عن الحقول المزروعة بالقمح في المحافظة، نستطيع استخدام هذه الخارطة لسحب عينة عشوائية من جميع الحقول وباحتمال يتناسب مع حجم الحقل (المساحة) X_j و $j=1,2,\ldots,N$. واضح أنه توجد علاقة قوية بين إنتاج الحقل من القمح Y_j و $j=1,2,\ldots,N$ وحجم الحقل X_j، ولنفرض أنه $Y_j=kX_j$ هذا يعني أن ناتج القمح يتناسب مع حجم الحقل. ولنفرض أننا سحبنا عينة حجمها n من الحقول ودون إرجاع وحصلنا على المشاهدات $y_1, y_2, \ldots, y_n$ وكذلك مساحة هذه الحقول $x_1, x_2, \ldots, x_n$، وكذلك نفرض أننا نعرف مجموع المساحة المزروعة بالقمح لهذا الموسم X .فإن تقدير مجموع إنتاج القمح (المجموع الكلي للمجتمع)

$$Y_{PPS}=\frac{X}{n}\sum_{j=1}^{n}\frac{y_j}{x_j}$$

واضح أن Y_{PPS} تقدير غير متحيز إلى $Y=\sum_{j=1}^{N}Y_j$ المجموع الكلي للمجتمع مع تباين

$$Var(Y_{PPS})=\frac{X}{n}\sum_{j=1}^{N}\left(Y_j/X_j\right)^2-\left(Y^2/n\right)$$

في حالة كون $Y_j=kX_j$ و $Y=kX$ نحصل على تقدير مثالي أي إن $Var(Y_{PPS})=0$ وهذا غير ممكن في الحياة العملية؛ لأنه إذا كان $Y=kx$، ونعرف X نستطيع معرفة Y دون الحاجة إلى التقدير.

إن استخدام العينة العشوائية ذات الاحتمالات المتناسبة مع حجم الوحدة، يصاحبه مشكلات عدة من أهمها كيف نختار وحدات العينة مع احتمال يتناسب مع أحجامها ودون إرجاع، وكذلك كيف نحصل على عينات S_i واحتمال سحب كل عينة π_i وأخيراً كيف يمكن اشتقاق الخصائص

الإحصائية للتقديرات والتي تعتمد على (s_i, π_i) ، يمكن التغلب على بعض هذه المشكلات أحياناً وذلك بسحب وحدات العينة مع الإرجاع. كما مر معنا أعلاه. يراجع Singh and Chaudhary (1986) لمزيد من المعلومات حول بعض طرق اختيار وحدات العينة ودون إرجاع من المجتمع وكذلك حول بعض التفصيلات الأخرى.

4.11 تقدير النسبة R

لنفرض أننا سحبنا عينة $(y_1, x_1), (y_2, x_2), \ldots, (y_n, x_n)$ دون إرجاع حيث إن احتمال الحصول على Y_j بتناسب مع X_j هذا يعني

$$P(y_j) = kX_j = X_j / \sum_{i=1}^{N} x_i$$

كذلك نفرض أننا نعرف $X = \sum_{i=1}^{N} x_i$ وأن $Y_i = \beta X_i$. نلاحظ المعاينة مع احتمالات متناسبة مع الحجم للمتغير x هي نفسها احتمالات متناسبة مع الحجم للمتغير y لأن

$$P(y_j) = kX_j = X_j / \sum_{i=1}^{N} x_i = Y_j / \sum_{i=1}^{N} y_i$$

لنأخذ التقدير

$$Y_{PPS} = \frac{1}{n} \sum_{j=1}^{n} y_j / p_j = \frac{X}{n} \sum_{i=1}^{n} y_i / x_i$$

إن $E(Y_{PPS}) = Y$ أي أنه تقدير غير متحيز مع تباين يساوي صفراً. نلاحظ أن Y_{PPS} مشابه إلى تقدير النسبة

$$R_1 = \frac{1}{n} \sum_{i=1}^{n} y_i / x_i$$

ولكن مع الفرق لإيجاد R_1 استخدمنا المعاينة العشوائية البسيطة متساوية الاحتمالات، أما Y_{PPS} استخدمنا المعاينة المختلفة الاحتمالات أو بالتحديد احتمال سحب الوحدات يتناسب مع حجمها. ولابد من الإشارة هنا أننا في الحياة الواقعية لا يمكن أن نحصل على $Y_i = \beta X_i$ ولكن قد نحصل على حالات قريبة من هذه العلاقة ولذلك يبقى مفيداً في الحياة العملية.

لنفرض أننا لدينا الأمثلة الآتية Y يمثل إنتاج القمح و X مجموع الدونمات المزروعة بالقمح، Y يمثل إنتاج شركات من بضاعة معينة و X مجموع العاملين في هذه الشركات، Y يمثل مصروف العائلات و X حجم العائلات. والسؤال الآن: كيف نختار وحدات عينة باحتمال يتناسب مع حجم الوحدات؟ إذا كان لدينا $X_1, X_2, ..., X_N$. سوف نتناول طريقتين .

1. الطريقة الأولى استخدام المجموع الجزئي

يمكن أن نجد المجموع الجزئي وهو

$$X_1,\ X_1+X_2,\ X_1+X_2+X_3,\ ...,\ \sum_{i=1}^{N} X_i$$

نقوم بسحب رقم عشوائي Z يقع بين 1 إلى $X = \sum_{i=1}^{N} X_i$ ونختار X_j عندما

$$\sum_{i=1}^{j-1} X_i < Z \leq \sum_{i=1}^{j} X_i$$

ولكن هذه الطريقة تستغرق وقتاً طويلاً ولا بد من توافر كشف بجميع $X_1, X_2, ..., X_N$.

2. الطريقة الثانية تسمى (Lahiri's Method)

لنفرض أننا عندنا فكرة عن أكبر قيمة من بين X_j و $j=1, 2, \ldots, N$ ولنعطها اسماً X_{max} ، نقوم الآن بسحب رقمين عشوائيين مستقلين عن بعضهما بعضا ، الأول عدد صحيح يقع بين (1 , N) والثاني يقع بين $(0, X_{max})$. لنفرض أنهما j و Z على التوالي، إذا كان $Z<X_j$ نأخذ X_j لتكون هي المشاهدة المطلوبة، خلاف ذلك نرفض الرقمين ونسحب رقمين جديدين. وهكذا. ولكن هذه الطريقة ستكون حساسة جداً للخطأ في قيمة X_{max} ؛ لأننا في الغالب لا نعرفه ونحاول تحديده بطريقة أو بأخرى، لمزيد من المعلومات يراجع Lahiri (1951).

5.11 العينة العشوائية المنتظمة

سوف نتناول هنا العينة المنتظمة مع كون احتمال سحب وحداتها يتناسب مع حجم الوحدة، والتي اقترحها Madow (1949)، يمكن تلخيص هذه الطريقة، بأن نقوم بترتيب وحدات المجتمع بصورة عشوائية، ومن ثم نقوم ببناء ما يسمى بالمجاميع التراكمية $T_i=\sum_{j=1}^{i} X_j$ و $i=1, 2, \ldots, N$ حيث إن X_j يمثل حجم الوحدة i. لاختيار عينة بحجم n وباحتمال متناسب مع حجم الوحدة، نقوم بسحب رقم عشوائي وليكن R يقع بين 1 إلى k و $k=\frac{T_N}{n}$ تكون الوحدة المعادلة للرقم $R+jk$ و $j=0, 1, \ldots, (n-1)$ في العينة.

مثال (1): لنفرض أن لدينا قرية تحتوي على 10 مناطق زراعية تحتوي على العدد الآتي من الحقول: 27 ,28 ,35 ,24 ,26 ,40 ,25 ,45 ,30 ,50 ونريد أن

نختار عينة عشوائية تتكون من 3 مناطق وبطريقة منتظمة مع احتمال يتناسب مع عدد الحقول في هذه المناطق.

الحل:

$$T_N = 330 \Rightarrow k = \frac{T_N}{n} = \frac{330}{3} = 110$$

لنفرض أن الرقم الذي اختير بطريقة عشوائية يقع بين 1 إلى 110 هو 51. إذن الوحدات المقابلة للمجاميع التراكمية 51 , 161 , 271 تكون في العينة المنتظمة مع احتمال يتناسب مع حجم الوحدات (المناطق). يمكن استخدام هذه الطريقة إذا كان k عدداً غير صحيح وذلك بأخذ أقرب عدد صحيح إلى T_N/n .

لقد اشتق Hartley and Rao (1962) تقديراً غير متحيز لتقدير المجموع الكلي وهو

$$Y_{HR} = \sum_{i=1}^{n} \frac{y_i}{\pi_i} = \sum_{i=1}^{n} \frac{y_i}{np_i}$$

ولكن المشكلة التي واجهتنا في تقدير التباين في العينة العشوائية المنتظمة والتي تكلمنا عنها في الفصل الثامن، تواجهنا هنا في تقدير تباين Y_{HR} ، ولقد اقترح Hartley and Rao تبايناً إلى Y_{HR} يمكن استخدامه إذا كان حجم العينة صغيراً قياساً إلى حجم المجتمع وهو:

$$Var(Y_{HR}) \approx \frac{1}{n} \sum_{i=1}^{N} (y_i/p_i - Y)^2 p_i \, [1 - (n-1) p_i]$$

ويمكن تقدير هذا التباين باستخدام

$$s^2(Y_{HR}) = \frac{1}{n^2(n-1)} \sum_{i=1}^{n} \sum_{j>i}^{n} [1 - n(p_i + p_j) + \sum_{j}^{N} p_j^2] (\frac{y_i}{p_i} - \frac{y_j}{p_j})^2$$

لتفاصيل ومعلومات أكثر عن الموضوع يراجع Hartley and Rao (1962).

6.11 العينة العنقودية بمرحلة واحدة

لقد مر معنا في الفصل التاسع في بعض الأحيان نستطيع أن نقلل التباين للتقدير عندما نقوم باختيار وحدات العينة باحتمال يتناسب مع حجم الوحدة، في الحقيقة أن العينة العنقودية غالباً ما تعطينا حالات مثالية لفكرة اختيار وحدات العينة باحتمال يتناسب مع حجم الوحدة، لأن عدد الوحدات في العنقود الواحد m_i في الغالب غير متساوية.

إن استخدام العينة العنقودية مع احتمال يتناسب مع حجم العنقود m_i يعطينا فائدة كبيرة في تقليل التباين خصوصاً إذا كان مجموع العناقيد y_i له علاقة قوية أو عالية مع عدد العناصر في العنقود. لنفرض $p_i = \frac{m_i}{M}$ احتمال أن الوحدة i تظهر في العينة، حيث إن $M = \sum_{i=1}^{N} m_i$ وN يمثل عدد العناقيد في المجتمع، لذلك فإن تقدير المجموع الكلي للمجتمع يكون

$$Y_{pps} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{p_i} = \frac{M}{n} \sum_{i=1}^{n} \frac{y_i}{m_i}$$

أما تقدير الوسط الحسابي للمجتمع فيكون

$$\bar{y}_{pps} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{m_i}$$

إن تقدير التباين إلى Y_{pps} و $\bar{y}_{pps}$ فهما على التوالي

$$s^2(Y_{pps}) = \frac{M^2}{n(n-1)} \sum_{i=1}^{n} (y_i/m_i - \bar{y}_{pps})^2$$

$$s^2(\bar{y}_{pps}) = \frac{1}{n(n-1)} \sum_{i=1}^{n} (y_i/m_i - \bar{y}_{pps})^2$$

7.11 العينة العنقودية بمرحلتين

إن عـدد الوحـدات للعناقيـد مختلـفٌ مـن عنقـود إلى آخـر أحيانـاً بشـكل كبير، لذلك فإن سحب العناقيد من المجتمع باحتمالات متناسبة مع حجم هذه العناقيد غالباً ما يفضل على سحبها باحتمالات متساويةً، والذي تكلمنا عنه في الفصل العاشر، بصورة عامة نستخدم احتمالات متناسبة مع حجم العنقود في المرحلة الأولى فقط، لأن العناصر داخل كل عنقود غالباً ما تكون متساوية الحجم. لذلك سوف نتكلم عن تقدير للوسط الحسابي، والمجموع الكلي للمجتمع يعتمد على احتمالات متناسبة مع الحجم في المرحلة الأولى فقط. لقد حصلنا على تقدير للوسط الحسابي في الفقرة أعلاه وهو عبارة عن

$$\bar{y}_{pps} = \frac{1}{n} \sum_{i=1}^{n} \frac{y_i}{m_i} = \frac{1}{n} \sum_{i=1}^{n} \bar{y}_i$$

و $\bar{y}_i$ وجدناه باستخدام جميع عناصر العنقود i. هنا نحسب $\bar{y}_i$ بالاعتماد على عينة عشوائية بسيطة من عناصر العنقود i، لذلك $\bar{y}_i$ هنا ما هو إلا تقدير إلى الوسط الحسابي للعنقود i. إن تقدير الوسط الحسابي في العينة العنقودية

بمرحلتين مع استخدام احتمالات متناسبة مع حجم العنقود في المرحلة الأولى سيكون

$$\bar{y}_{pps} = \frac{1}{n}\sum_{i=1}^{n}\bar{y}_i$$

مع تقدير التباين

$$s^2(\bar{y}_{pps}) = \frac{1}{n(n-1)}\sum_{i=1}^{n}(\bar{y}_i - \bar{y}_{pps})^2$$

وكذلك فإن تقدير المجموع الكلي سيكون

$$Y_{pps} = \frac{M}{n}\sum_{i=1}^{n}\bar{y}_i$$

مع تقدير للتباين

$$s^2(Y_{pps}) = \frac{M^2}{n(n-1)}\sum_{i=1}^{n}(\bar{y}_i - \bar{y}_{pps})^2$$

References

1. Hartley, H. O. and Rao, J. N. K. (1962). Sampling without Replacement, J. Amer. Statist. Assoc., 61, 739–748.
2. Horvitz, D. G. and Thompson, D. J. (1952). A Generalization of Sampling without Replacement, *J. Amer. Statist. Assoc.* 47, 663–685.
3. Lahiri, D. B. (1951). A Method of Sample Selection Providing Unbiased Ratio Estimates Bull. Int. Statist., 33, 133, 140.
4. Lahiri, D. B. (1951). A Method of Sample Selection Providing Unbiased Ratio Estimates Bull. Int. Statist., 33, 133, 140.
5. Madow, N. G. (1949). On Theory of Systematic Sampling Ann. Math. Statist., 20, 333–354.
6. Ray, D. (1956). Some Estimates in Sampling with Varying Probabilities without Replacement. *J. Amer. Statist. Assoc.* 51, 269–284.
7. Samford, M. R. (1967). On Sampling without Replacement with Unequal Probabilities of Selection, *Biometrika* 54, 499–513.
8. Scheaffer, R.L., Mendenhall, W. and Ott, L. (1996). Elementary Sampling Survey, 5th ed., Duxbury, New York.
9. Singh, D. and F. S. Chaudhary (1986). Theory and Analysis of Sample Survey Designs, John Wiley & Sons, New York.
10. Thompson, S. K (2002). Sampling, 2end Wiley, New York.

الفصل الثاني عشر

المعاينة المزدوجة
Double sampling

1.12 مقدمة

لقد سبق أن تكلمنا في الفصل السابع عن عدة طرق للمعاينة تعتمد على المعلومات المساعدة أو الإضافية أو ما يسمى بالمتغير المساعد x ،ولقد لاحظنا أن التقدير باستخدام النسبة R أو خط انحدار y على x أو الفرق يتطلب معرفة الوسط الحسابي للمجتمع للمتغير x ،وإذا كنا نرغب في تقسيم المجتمع إلى طبقات بالاعتماد على المتغير x فإن ذلك يتطلب معرفة التوزيع التكراري للمتغير المساعد x.

عندما لا تكون المعلومات الخاصة بالمتغير المساعد x متوافرة، ولكن في بعض الأحيان غير مكلفة أن نسحب عينة عشوائية كبيرة من المجتمع للمتغير X وحده ونجمع معلومات حوله. إن الهدف من هذه العينة الحصول على تقدير جيد إلى الوسط الحسابي للمجتمع أو التوزيع التكراري للمتغير المساعد X في كثير من المسوحات التي يكون الهدف منها تقدير معلمات المجتمع للمتغير y وحده، يمكن أن يكون مجديًا ونافعًا أن يخصص بعض الموارد المخصصة للمسح لسحب عينة أولية للحصول على معلومات حول المتغير المساعد x على الرغم من أن ذلك يعني أن حجم العينة لتقدير المتغير y سوف يتقلص، إن هذه الطريقة تعرف باسم المعاينة المزدوجة (Double sampling).

إن هذه الطريقة للمعاينة مفضلة، إذا كانت الدقة التي نحصل عليها من خلال استخدام السبة، أو خط الانحدار، أو استخدام الطبقية تكون أكبر فيما لو استخدمنا جميع الموارد لسحب عينة بحجم معين للمتغير الأساس y وحده.

212 تقدير النسبة باستخدام العينة المزدوجة

الرموز والمصطلحات الآتية سوف نستخدمها في هذا الفصل

y_i : قيمة المتغير الأساسي، أو الصفة تحت الدراسة للوحدة i في المجتمع.

x_i : قيمة المتغير المساعد، أو الصفة المساعدة في نفس الوحدة i في المجتمع.

n' : عدد الوحدات التي جرى سحبها في العينة الأولى.

n : عدد الوحدات التي جرى سحبها في العينة الثانية.

ستجري مشاهدة كلا المتغيرين y وx لكل وحدة من وحدات العينة الثانية، أما في العينة الأولى فسوف تتم مشاهدة المتغير x فقط.
إذا كان المتغيرين y وx مرتبطين بعلاقة خطية قوية بحيث إذا كانت $x_i = 0$ فإن قيمة $y_i = 0$ ، لذا فإن تقدير النسبة مع العينة المزدوجة سيؤدي إلى تقدير أفضل لمعدل أو مجموع المجتمع للمتغير الأساسي y.

لنفترض أنه من مجتمع حجمه N جرى سحب عينة عشوائية مع عدم الإرجاع وبحجم n' من الوحدات وتمت مشاهدة المتغير x، كذلك تم سحب عينة عشوائية مع عدم الإرجاع وبحجم n من بين وحدات العينة الأولى n' وجرى مشاهدة المتغير y بالإضافة إلى المتغير x.

نستطيع أن نجد من العينة الصغيرة نسبة العينة وهو

$$r = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i}$$

حيث إن المجموع أخذ لجميع عناصر العينة الثانية.

نستطيع من العينة الكاملة التي تم مشاهدة المتغير x فيها أن نقدر مجموع المجتمع لهذا المتغير وهو

$$\hat{t}_x = \frac{N}{n'} \sum_{i=1}^{n'} x_i$$

ويمكننا أن نجد القيمة التقريبية لتباين $\hat{t}_x$

$$Var(\hat{t}_x) \approx N(N-n')\frac{s^2}{n'} + N^2\left(\frac{n'-n}{n'}\right)\frac{s_r^2}{n}$$

حيث أن σ^2 عبارة عن تباين المجتمع للمتغير y و σ_r^2 يمثل تباين المجتمع حول خط النسبة ويمكن أن نعرفه بما يأتي :

$$\sigma_r^2 = \frac{1}{N-1} \sum_{i=1}^{N} (y_i - Rx_i)^2$$

بحيث أن $R = \tau / \tau_x$ وهو نسبة المجتمع. ويمكننا أن نحصل على تقدير للتباين $var(\hat{t}_x)$ وهو

$$s^2(\hat{t}_x) \approx N(N-n')\frac{s^2}{n'} + N^2\left(\frac{n'-n}{n'n(n-1)}\right) \sum_{i=1}^{n} (y_i - rx_i)^2$$

حيث أن s^2 هو تباين العينة للمتغير y.

مثال 1: تمت مشاهدة 240 غزال في مسح من الجو لعينة تتكون من 20 وحدة، علماً أن المجتمع يتكون من 100 وحدة متساوية المساحة تقريباً، ولقد

تم إرسال مجموعة من العادين؛ لكي يقوموا بمشاهدة 5 وحدة من بين الوحدات العشرين التي تم مسحها من الجو؛ لمشاهدتها بصورة تامة على الأرض، ولقد تمت مشاهدة 56 غزال في هذه الوحدات الخمس من الجو، ولكن عندما تمت مشاهدتها على الأرض تم رصد 70 غزال، وهذا يمثل الرقم الحقيقي لعدد الغزلان في هذه القطع الخمس. قدّر عدد الغزلان في منطقة الدراسة.

الحل:

$$r = \frac{\sum_{i=1}^{5} y_i}{\sum_{i=1}^{5} x_i} = \frac{70}{56} = 1.25$$

باستخدام العينة الكاملة $n' = 20$ نستطيع أن نقدر المجموع الكلي للمجتمع

$$\hat{t}_x = \frac{N}{n'} \sum_{i=1}^{n'} x_i = \frac{100}{20} \sum_{i=1}^{20} x_i = \frac{100}{20}(240) = 1200$$

الآن نستطيع أن نقدر عدد الغزلان باستخدام تقدير النسبة والمعتمد على العينة المزدوجة وهو

$$\hat{t}_y = r\,\hat{t}_x = 1.25(1200) = 1500$$

أي أننا نقدر مجموع الغزلان في منطقة الدراسة بنحو 1500 غزال.

3.12 توزيع العينة في المعاينة المزدوجة لتقدير النسبة

سيكون تقدير النسبة باستخدام العينة المزدوجة أكثر فاعلية إذا كان المتغيران y وx مرتبطين بعلاقة خطية قوية، بحيث إن قيمة $y=0$ إذا كانت $x=0$، وع كون تكلفة قياس المتغير x أرخص من y. إن القيمة المثلى لنسبة

حجم العينة الجزئية n إلى حجم العينة الكلية n' يعتمد على تكاليف مشاهدة أو قياس المتغيرين y وx وكذلك على قوة العلاقة بينهم.

لنفرض أن تكاليف مشاهدة وحدة واحدة للمتغير x هي c' وتكاليف مشاهدة المتغير y هي c، لذا فإن مجموع التكاليف C ستكون

$$C = c'n' + cn$$

إذا كانت التكاليف C ثابتة فإن الحد الأدنى لقيمة التباين للمتغير □ يمكن إيجادها باستخدام العلاقة الآتية:

$$\frac{n}{n'} = \sqrt{\frac{c'}{c}\left(\frac{\sigma_r^2}{\sigma^2 - \sigma_r^2}\right)}$$

4.12 المعاينة المزدوجة للتقسيم المجتمع إلى طبقات

لا يمكن تصنيف الوحدات إلى طبقات في بعض الحالات إلا بعد أن نقوم بعملية سحب الوحدة، على سبيل المثال عندما نقوم بسحب عينة من الأشخاص ومن ثم الاتصال بهم هاتفياً لجمع معلومات منهم، لا يمكن تصنيفهم حسب الجنس أو المستوى التعليمي أو العمر أو نوع العمل إلا بعد الاتصال بهم هاتفياً. إن طريقة تقسيم المجتمع إلى طبقات بعد سحب العينة والتي سبق وأن تحدثنا عنها في الفصل السادس يمكن أن تكون مفيدة في مثل هذه الحالات إذا كان وزن الطبقة $W_i = N_i/N$ معلوماً لكل طبقة. إذا كان وزن الطبقة غير معلوم فيمكن استخدام المعاينة المزدوجة مع عينة أولية كبيرة يمكن استخدامها لتصنيف الوحدات إلى طبقات، ومن ثم نقوم بسحب عينة طبقية من العينة الأولية. يمكن أن تكون العينة المزدوجة مفيدة إذا كانت الوحدات في كل عينة متماثلة وإذا كان مشاهدة المتغير x أسهل وأقل تكلفة من المتغير y.

نقوم بسحب عينة عشوائية أولية بحجم n' وحدة من المجتمع ذي الحجم N، ومن ثم نقوم بتصنيف هذه الوحدات إلى طبقات في كل طبقة n'_i وحدة جرى مشاهدتها في الطبقة i و $i = 1, 2, \ldots, K$. نقوم بتقدير وزن الطبقة في المجتمع W_i باستخدام وزن الطبقة في العينة $w_i = \frac{n'_i}{n'}$ ، ومن ثم نقوم بسحب عينة طبقية من العينة الأولية بحيث نسحب n_i وحدة من الوحدات n'_i في الطبقة i، وأخيراً نقوم بمشاهدة وتسجيل جميع قيم المتغير y لكل وحدة في العينة الثانية. يمكن أن نعرف الوسط الحسابي للعينة الثانية للطبقة i

$$\bar{y}_i = \frac{1}{n_i}\sum_{i=1}^{n_i} y_i$$

وتقدير الوسط الحسابي للمجتمع يكون

$$\bar{y}_d = \sum_{i=1}^{K} w_i \bar{y}_i$$

التقدير $\bar{y}_d$ سيكون تقدير غير متحيز إلى الوسط الحسابي للمجتمع مع تباين

$$var(\bar{y}_d) = \frac{N-n'}{N}\frac{\sigma^2}{n'} + E\sum_{i=1}^{K}\left[\left(\frac{n'_i}{n'}\right)^2 \left(\frac{n'_i-n_i}{n'_i}\right)\frac{\sigma^2_{i(s_1)}}{n_i}\right]$$

حيث أن σ^2 يمثل التباين الكلي للمجتمع و $\sigma^2_{i(s_1)}$ يمثل تباين المجتمع في الطبقة i للوجه الأول للعينة s_1 . يمكن أن نحصل على تقدير غير متحيز لتباين $\bar{y}_d$ وهو

$$s^2(\bar{y}_d) = \frac{N-1}{N}\sum_{i=1}^{K}\left(\frac{n'_i-1}{n'-1} - \frac{n_i-1}{N-1}\right)\frac{w_i s_i^2}{n_i} + \frac{N-n'}{N(n'-1)}\sum_{i=1}^{K} w_i(\bar{y}_i - \bar{y}_d)^2$$

إذا جرى تحديد النسبة n_i/n'_i وجرى استبعاد العينات بحجم $n'_i = 0$ فإن تباين $\bar{y}_d$ يكون تقريباً

$$var(\bar{y}_d) = \frac{N-n'}{Nn'}\sigma^2 + \sum_{i=1}^{K}\frac{W_i\sigma_i^2}{n'}\left(\frac{n'_i}{n_i}-1\right)$$

ويبقي تقدير التباين كما هو لهذه الحالة. يراجع Thompson (2002) لمزيد من المعلومات.

5.12 التقديرات باستخدام الانحدار في العينة المزدوجة

من بعض تطبيقات المعاينة المزدوجة يمكن استخدام المتغير المساعد تقدير الوسط الحسابي $\bar{Y}$ للمجتمع باستخدام الانحدار. لنفرض أنه جرى سحب عينة عشوائية بحجم n' من الوحدات وتمت مشاهدة المتغير المساعد x، كذلك تم سحب عينة عشوائية بحجم n وجرى مشاهدة المتغير الأساسي y بالإضافة إلى المتغير المساعد x. لذا فإن تقدير $\bar{Y}$ باستخدام انحدار y على x يمكن تعريفه

$$\bar{y}_{lr} = \bar{y} + b(\bar{x}' - \bar{x})$$

حيث $\bar{x}'$ و $\bar{x}$ هما متوسطا المتغير المساعد x في العينتين الأولى والثانية و b معامل انحدار المربعات الصغرى للمتغير y على x محسوباً من العينة الثانية. وإذا لم نفترض بوجد علاقة خطية بين المتغيرين في المجتمع فإن $\bar{y}_{lr}$ سيكون منحازاً. وإذا افترضنا يمكننا إهمال $1/n$ و $1/n'$ ويمكننا إعطاء تقدير تقريبي لتباين $\bar{y}_{lr}$

$$var(\bar{y}_{lr}) \approx \frac{S_y^2(1-\rho^2)}{n} + \frac{\rho^2 S_y^2}{n'} - \frac{S_y^2}{N}$$

حيث S_y^2 يمثل تباين المجتمع للمتغير y و ρ معامل الارتباط للمجتمع بين المتغيرين y و x.

أما تقدير التباين $1/n$ و $1/n'$ فسيكون

$$s^2(\bar{y}_{lr}) = \frac{s_{yx}^2}{n} + \frac{s_y^2 - s_{yx}^2}{n'} - \frac{s_y^2}{N}$$

حيث

$$s_{yx}^2 = \frac{1}{n-2}\left[\sum_{i=1}^{n}(y_i - \bar{y})^2 - b^2 \sum_{i=1}^{n}(x_i - \bar{x})^2\right]$$

و

$$s_y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2$$

أما إذا لم يكن بالإمكان إهمال $1/n$ لصغر حجم العينة، فإن التقدير المقترح لتقدير التباين سيكون

$$s^2(\bar{y}_{lr}) = s_{yx}^2\left[\frac{1}{n} + \frac{(\bar{x}' - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right] + \frac{s_y^2 - s_{yx}^2}{n'} - \frac{s_y^2}{N}$$

لمزيد من المعلومات يراجع (1977) Cochran، أو ترجمة أنيس كنجو(1995).

6.12 تكرار المعاينة لنفس المجتمع

لقد أصبح الاعتماد على العينة لجمع المعلومات حول مجتمعات الدراسة المختلفة وخصوصاً الإنسانية منها والتي هي لمدد زمنية متقاربة، من الأمور المتعارف عليها بل أحياناً أصبحت ضرورةً ملحة؛ وذلك لكون المجتمعات متحركة ومتبدلة بصورة كبيرة، وأصبحت المسوحات التي تجرى كل عشر سنوات مثل المسوحات السكانية قليلة الفائدة بعد سنة أو سنتين؛ وذلك للتغيرات السريعة والكبيرة التي تحصل بسبب تبدل آراء الناس حول مختلف الموضوعات؛ وذلك للتطور العلمي والتكنولوجي الذي يشهده العالم هذه السنوات وخصوصاً العشرة سنوات الأخيرة.

عندما تجرى المسوحات بواسطة العينة ولُمددٍ متقاربة يكون الباحث في موضع مثالي ليقوم بتقديرات واقعية للتكاليف والتباينات ويمكنه من تطبيق الطريقة التي تمكنه من الوصل إلى فاعلية مثلى للمعاينة. من الأسئلة المهمة هي ما هي الطريقة التي تتم بها تغير العينة مع مرور الزمن؟ وما هي المدة الزمنية اللازمة للتغير؟ هنالك اعتبارات كثيرة قد تؤثر على القرار. ربما لأنَّ

بعض الناس لا يرغبون بإعطاء المعلومات نفسها مرة بعد أخرى، وربما يكون بعض الناس متأثرين ببعض المعلومات المسبقة حول المسح قد تمنعهم من إعطاء معلومات دقيقة في المقابلة الأولى، في بعض الأحيان التعاون يكون أفضل في المقابلة الثانية منه في الأولى والمعلومات التي تعطى تكون أكثر دقة خصوصاً إذا كانت المعلومات المطلوبة تقنية أو سرية أو شخصية.

إذا حصلنا على بيانات من عينات متتابعة، فهناك ثلاثة أنواع من التقديرات ربما يفكر الباحث القيام بها وهي:

1. التغير في الوسط الحسابي للمجتمع $\bar{Y}$ من مناسبة إلى المناسبة التي تليها.
2. معدل الوسط الحسابي للمجتمع $\bar{Y}$ لجميع المناسبات.
3. المعدل للوسط الحسابي للمجتمع $\bar{Y}$ في المناسبات الأحدث.

في كثير من المسوحات بالعينة يكون التركيز على المعدل للمناسبات الأحدث، خصوصاً إذا كان التغير في المجتمع يكون سريعاً وكبيراً من مدة زمنية لأخرى، ولكن إذا كان التغير بطيئاً فإنَّ المعدل لجميع المناسبات هو الأفضل، أما التغير في الوسط الحسابي للمجتمع من مناسبة لأخرى فيستعمل في الغالب لمعرفة التغيرات التي قد تحصل على الوسط الحسابي إذا حدثت ظروف نتوقع أنها تؤثر على الوسط الحسابي للمجتمع بشكل ملحوظ. على سبيل المثال رفع رواتب العاملين بالدولة وتأثيره على ارتفاع الأسعار أو زيادة الاستثمار.

لنفرض أننا لدينا الحرية في تغير تركيب العينة أو الاحتفاظ به، وأن حجم العينة الكلي سيبقي نفسه في جميع المناسبات، إذا رغبنا الحصول على أعلى دقة فيمكننا وضع العبارات الآتية حول سياسة الاستبدال:

1. من الأفضل أن نحافظ على نفس العينة خلال جميع المناسبات إذا رغبنا في تقدير التغير.
2. إذا رغبنا في تقدير المتوسط لجميع المناسبات من الأفضل سحب عينة جديدة في كل مناسبة.
3. إذا كنا نركز على التقديرات الراهنة، فإننا نحصل على الدقة نفسها سواء احتفظنا بالعينة نفسها أو بدّلناها في كل مناسبة، وقد يكون من الأفضل تغير جزء من العينة في كل مناسبة.

تمارين

1. في مسح جوي لأربع قطع جرى سحبها من بين عشر قطع في منطقة الدراسة باستخدام العينة العشوائية البسيطة، لوحظ أنَّ عدد طيور البط التي جرى مشاهدتها في هذه القطع الأربعة هي على النحو الآتي: 44, 55, 4, 16. ولكن عند فحص الصور الفوتوغرافية للقطعة الأولى والثالثة (التي جرى سحبهما باستخدام العينة العشوائية البسيطة بصورة دقيقة) تبين أن عدد طيور البط في هاتين القطعتين هو 56 و6. قدر مجموع البط في منطقة الدراسة باستخدام تقدير النسبة، ومن ثم أوجد تقدير التباين لهذا التقدير.

2. لتقدير المصروفات الشهرية للمصروفات على العلاج الطبي في مدينة تحتوي على 5000 عائلة، جرى سحب عينة عشوائية تحتوي على500 عائلة، فوجد أن هنالك 336 عائلة لديها أطفال و164 دون أطفال، ومن ثم جرى سحب عينة جزئية طبقية تحتوى على 112 عائلة لديها أطفال و 41 لا يوجد لديها أطفال، ولقد جرى جمع معلومات حول المصروفات الشهرية لأفراد العينة الجزئية وحصلنا على المعلومات الآتية:

	الوسط الحسابي	الانحراف المعياري	حجم العينة
دون أطفال	$280	$160	41
مع أطفال	$110	$60	112

قدر معدل المصروفات الطبية الشهرية للعائلة، ومن ثم قدر تباين التقدير.

3. في أحد تطبيقات المعاينة المزدوجة مع الانحدار كان حجم العينة الصغيرة 87 وحجم العينة الكبيرة 300. النتائج الآتية تتعلق بالعينة الصغيرة

$$\sum_{i=1}^{87}(y_i-\bar{y})(x_i-\bar{x}) = 5114 \text{ و } \sum_{i=1}^{87}(x_i-\bar{x})^2 = 3248 \text{ و } \sum_{i=1}^{87}(y_i-\bar{y})^2 = 17283$$

أوجد تقدير الخطأ المعياري لتقدير الانحدار .

المراجع العربية

1. وليم كوكوران – ترجمة أنيس كنجو ،(1995) تقنية المعاينة الإحصائية ، جامعة الملك سعود ، الرياض.

References

1. Cochran, W. G. (1961). Comparison of Methods for Determined Stratum Boundaries. *Bull. Inst. Stat. Inst.*, 38, 345–358.
2. Cochran, W. G. (1977). Sampling Technique, 3rd Ed. Wiley, New York.
3. Rao, J. N. K. (1973). On Double Sampling for Stratification and Analytical Surveys. *Biometrika*, 60, 125–133.
4. Sämdal, C. E. and Swensson, B. (1987). A General View of Estimation of Two Phases of Selection with Applications to Two–phase Sampling and Nonresponse. *International Statistical Review*, 55,279–294.
5. Schafer, J. L. (1997). Analysis of Incomplete Multivariate Data, London Chapman & Hall.
6. Singh, D. and Chaudhary, F. S. (1986). Theory and Analysis of Sample Survey Designs, New Delhi, Wiley Eastern.
7. Thompson, S. K. (2002). Sampling, 2end Wiley, New York.

الفصل الثالث عشر

تقدير حجم المجتمع
Estimation of the Population Size

1.13 مقدمة

لقد تناولنا في الفصول السابقة تقدير الوسط الحسابي والمجموع الكلي والنسبة للمجتمع مع افتراض أن حجم المجتمع N معلوم، وإذا كان غير معلوم فعلينا أن نقوم بإهمال معامل التصحيح ونفترض أن حجم المجتمع كبير جداً، لكن في كثير من الأحيان يكون من الأهمية بمكان تقدير حجم المجتمع، فعلى سبيل المثال تقدير عدد الأسماك الموجودة في خليج العقبة أو الخليج العربي، أو تقدير عدد المتظاهرين الذين خرجوا في مظاهرة معينة، أو تقدير عدد الحيوانات البرية الموجودة في إحدى الصحاري ... إلخ.

هنالك أكثر من طريقة يمكن استخدامها لتقدير حجم المجتمع سوف نتناول ثلاثة منها بالتفصيل في هذا الفصل، وهي: الطريق المباشرة، والطريقة المعكوسة، والطريقة التي تعتمد على تقدير كثافة الوحدات في المجتمع.

2.13 تقدير حجم المجتمع باستخدام المعاينة المباشرة

نقوم في هذه الطريقة بسحب عينة عشوائية من أفراد المجتمع (الحيواني)، ونضع علامات مميزة على جميع الوحدات (الحيوانات) التي سحبت في العينة، ونقوم بإعادتها إلى المجتمع، وتترك لها الفرصة الكافية لكي تختلط بالمجتمع مرة أخرى، ثم نقوم في وقت لاحق بسحب عينة ثانية محددة الحجم من نفس

المجتمع، ونقوم بحصر عدد الوحدات (الحيوانات) التي تحمل العلامة المميزة والتي وضعت على الوحدات (الحيوانات) التي سحبت في العينة الأولى. إذا كانت N تمثل عدد أفراد المجتمع وr عدد الوحدات التي تحمل العلامة المميزة و$P = r / N$ يمثل نسبة الوحدات التي تحمل العلامة المميزة .لذا فإننا نستطيع تقدير N باستخدام $\hat{p} = \frac{s}{n}$ التي تمثل نسبة الوحدات والتي تحمل العلامة المميزة في العينة الثانية حيث إن s تمثل عدد الوحدات التي تحمل العلامة المميزة في العينة الثانية وn يمثل حجم العينة الثانية.

$$\hat{N} = \frac{r}{\hat{p}} = \frac{rn}{s}$$

مع تقدير للتباين

$$\hat{V}(\hat{N}) = s^2(\hat{N}) = \frac{r^2 n(n-s)}{s^3}, \quad s > 0$$

نلاحظ أن $\hat{N}$ تقدير متحيز إلى N حيث إن

$$E(\hat{N}) = N + \frac{N(N-r)}{nr}$$

إذا كان حجم r وn كبيرًا فإن مقدار التحيز $\frac{N(N-r)}{nr}$ سيكون صغيراً ويمكن إهماله، كذلك يمكن أن نلاحظ أن $\hat{N}$ يبالغ بزيادة تقديرN. ولمزيد من المعلومات، ولتقديرات أخرى لـ N يراجع Chapman (1952).

مثال (1): رغب مدير دائرة الحيوانات البرية في إحدى المحافظات بتقدير عدد الغزلان في الصحراء التابعة لمحافظته فنُصح بأن يسحب عينة عشوائية أولية بحجم $r = 100$ من الغزلان ويقوم بوضع علامة مميزة على جميع هذه الغزلان ومن ثم إرجاعها إلى الصحراء التي كانت تعيش فيها لمدة شهر، ومن ثم يقوم بسحب عينة ثانية بحجم $n = 50$ غزال، فقام بما نُصِح بعمله، ومن ثم قام

بحصر عدد الغزلان التي تحمل الصفة المميزة في العينة الثانية فوجدها $s = 10$. قَدِّر عدد الغزلان في هذه الصحراء، وأوجد حدًا أعلى لمقدار الخطأ لهذا التقدير.

الحل: يمكن تقدير عدد الغزلان باستخدام

$$\hat{N} = \frac{nr}{s} = \frac{100(50)}{10} = 500$$

يمكن إيجاد الحد الأعلى لمقدار الخطأ كما يأتي

$$2\sqrt{s^2(\hat{N})} = 2\sqrt{\frac{r^2 n(n-s)}{s^3}} = 2\sqrt{\frac{(100)^2 50(50-10)}{(10)^3}} = 282.84$$

لذلك فإننا نقدر عدد الغزلان بنحو 500 غزال وبحد أعلى للخطأ مقداره 282.84.

3.13 تقدير حجم المجتمع باستخدام المعاينة المعكوسة

هذه الطريقة مماثلة للطريق الأولى (العينة المباشرة)، ولكن هنا العينة الثانية لا يكون حجمها ثابتاً، بل نسحب وحدات من المجتمع إلى أن تصل إلى عدد معين (s) من الوحدات التي تحمل العلامة المميزة. وباستخدام هذه الطريقة نستطيع تقدير حجم المجتمع N باستخدام

$$\hat{N} = \frac{r}{\hat{p}} = \frac{rn}{s}$$

ولكن s هنا ثابتة وn متغيرة على خلاف ما كان عليه الحال في طريقة العينة المباشرة، وأما تقدير التباين إلى $\hat{N}$ فيكون

$$s^2(\hat{N}) = \frac{r^2 n(n-s)}{s^2(s+1)}$$

لابد من ملاحظة أن الشرط $s > 0$ لن يكون مشكلة في هذه الطريقة لأن s يجري تثبيتها من قبل الباحث على خلاف الطريقة السابقة، ولابد من الإشارة هنا إلى أن التقدير $\hat{N}$ وهو تقدير غير متحيز إلى N على خلاف الطريقة السابقة، وكذلك تقدير التباين فهو تقدير غير متحيز إلى التباين الحقيقي إلى $\hat{N}$.

مثال (2): لنفرض أننا في المثال السابق حددنا عدد الغزلان التي نريد أن نحصل عليها في العينة الثانية لتكون $s = 10$ وقمنا بسحب غزلان من المجتمع وبعد سحب 80 غزال حصلنا على 10 غزلان تحمل العلامة المتميزة، أي أن حجم العينة الثانية $n = 80$. قَدَّر عدد الغزلان في هذه الصحراء، وأوجد حدًا أعلى لمقدار الخطأ لهذا التقدير.

الحل: يمكن تقدير عدد الغزلان باستخدام

$$\hat{N} = \frac{nr}{s} = \frac{100(80)}{10} = 800$$

يمكن إيجاد الحد الأعلى لمقدار الخطأ كما يأتي

$$2\sqrt{s^2(\hat{N})} = 2\sqrt{\frac{r^2 n(n-s)}{s^2(s+1)}} = 2\sqrt{\frac{(100)^2 80(80-5)}{(10)^2(10+1)}} = 467.1$$

لذلك فإننا نقدر عدد الغزلان بنحو 800 غزال وبحد أعلى للخطأ مقداره 467.1.

4.13 تقدير حجم العينة باستخدام المعاينة المباشرة والمعكوسة

لقد تكلمنا أعلاه عن استخدام العينة المباشرة والمعكوسة لتقدير حجم المجتمع، وربما يتساءل القارئ أيهما أفضل أو أكثر دقة في تقدير حجم المجتمع؟ إن العينة المعكوسة تعطي معلومات أكثر دقة في تقدير حجم المجتمع من العينة المباشرة إذا كان حجم العينة الثانية n الذي يتطلب لإعادة إمساك s من الوحدات التي تحمل العلامة المميزة صغيراً بالمقارنة إلى حجم

المجتمع. ولكن إذا لم تكن هنالك أي معلومات متوافرة عن حجم المجتمع N فإن سوء اختيار أو تحديد حجم العينة الأولية r سيؤدي إلى أن يكون حجم العينة الثانية n كبيرًا جداً إذا تم استخدام المعاينة المعكوسة. على سبيل المثال إذا كان حجم المجتمع N=10000 وحدة وحجم العينة العشوائية الأولية r=50، فإن ذلك يتطلب أن يكون حجم العينة الثانية n كبيراً جداً للحصول على s=10 وحدة تحمل العلامة المميزة في العينة الثانية.

أولاً: تحديد حجم العينة الأولى r والثانية n للمعاينة المباشرة

يمكن استخدام جدول (1) لتحديد حجم العينيتين r وn المطلوبتين لتقدير حجم المجتمع N وبحد أعلى ثابت لمقدار خطأ التقدير. ولكن استخدام هذا الجدول يتطلب بعض المعلومات السابقة عن حجم المجتمع N، ولابد من الإشارة إلى أن محتويات الجدول رقم (1) هي $V(\hat{N})/N$ للمعاينة المباشرة. إذا كانت لدينا معلومات تقريبية عن حجم المجتمع N، يمكن أن نحدد تباين التقدير إذا كانت قيمتا r وn ثابتتين. يحتوي الجدول رقم (1) على قيمتي r وn من خلال النسب $p_1 = r/N$ و $p_2 = n/N$. لمزيد من المعلومات يراجع (2005) .Scheaffer et al.

جدول (1): قيم $V(\hat{N})/N$ للمعاينة المباشرة

	$p_1 = r/N$					
$p_2 = n/N$	0.001	0.01	0.1	0.25	0.5	1.0
0.001	999,000	99,000	9000	3000	1000	0
0.01	99,000	9,900	900	300	100	0
0.1	9,000	990	90	30	10	0
0.25	3,996	396	36	12	4	0
0.5	1,998	198	18	6	2	0
1.0	999	99	9	3	1	0

مثال (3):يعتقد مدير دائرة الحيوانات البرية في المثال رقم (1) أنَّ عدد الغزلان في الصحراء التابعة لمحافظته لهذا العام هو تقريباً نفس العدد في السنة الماضية وهو بين 500 إلى 700 غزال. أَوْجد حد الخطأ في التقدير المصاحب للنسب

$p_1 = 0.25$ و $p_2 = 0.1$. حدد حجم العينيتين r وn المطلوبتين إذا رغبت دائرة الحيوانات البرية استخدامهما لإجراء مسح جديد لتقدير حجم المجتمع N.

الحل: سوف نأخذ الحد الأعلى لعدد الغزلان وهو 700 حتى نحصل على تقدير محافظ للتباين $V(\hat{N})$. نستخدم الجدول رقم (1) مع النسب $p_1 = 0.25$ و $p_2 = 0.1$ لنحصل على

$$\frac{V(\hat{N})}{N} = 30$$

نأخذ N=700 للحصول على

$$V(\hat{N}) = 700(30) = 21\,000$$
$$\sqrt{V(\hat{N})} = \sqrt{21\,000} = 144.914$$

وعليه سيكون حد الخطاء في التقدير هو

$$2\sqrt{V(\hat{N})} = 2(144.914) = 289.828$$

إذا رغبت دائرة الحيوانات البرية بإجراء مسح بالعينة جديد فإن حجم العينة r يجب أن يكون

$$r = p_1 N = 0.25(700) = 175$$

أما حجم العينة n فيجب أن يكون

$$n = p_1 N = 0.1\,(700) = 70$$

ثانياً: تحديد حجم العينة الأولى r والثانية n للمعاينة المعكوسة

يمكننا أن نوجد التباين $V(\hat{N})$ للعينة المعكوسة بنفس الطريقة التي استخدمناها في العينة المباشرة أعلاه وباستخدام جدول رقم (2). ولابد من الإشارة إلى أن محتويات الجدول رقم (2) هي $V(\hat{N})/N$ للمعاينة المعكوسة. ويمكننا أن نستخدم الجدول رقم (2) وذلك باستخدام النسب $p_1 = r/N$ و $p_2 = n/N$. ولمزيد من المعلومات يراجع Scheaffer et al (2005).

جدول (2): قيم $V(\hat{N})/N$ للمعاينة المعكوسة

$p_2 = n/N$	$p_1 = r/N$					
	0.001	0.01	0.1	0.25	0.5	1.0
0.001	999	990	900	750	550	0
0.01		99	90	75	50	0
0.1			9	7.5	5	0
0.25				3	2	0
0.5					1	0
1.0						0

5.13 تقدير كثافة وحجم المجتمع باستخدام معاينة المربعات

إن تقدير عدد الوحدات في منطقة معينة (المجتمع) يمكن تحقيقه أولاً بتقدير عدد الوحدات لكل وحدة مساحة (أي كثافة الوحدات) ومن ثم ضربه بمساحة المنطقة (المجتمع) تحت الدراسة. على سبيل المثال إذا كان كاتب الطابعة يقوم بالمعدل بطباعة 3 كلمات خاطئة في الصفحة الواحدة، فإنه سيطبع تقريباً 600 كلمة خاطئة عند طباعته لكتاب يحتوى على 200 صفحة. سوف نناقش تقدير الكثافة وعدد الوحدات في منطقة معينة، ولكن هذه الطريقة لا تقتصر فقط على المساحة لأنه يمكن تطبيقها على الحجم مثل عدد البكتريا في سائل محدد الحجم، أو عدد المكالمات الهاتفية التي تتلقاها إلى دائرة معينة خلال مدة زمنية ثابتة، أو عدد الزائرين لموقع على الشبكة العنكبوتية (الإنترنت) خلال مدة زمنية محددة. ولكننا سنستخدم المساحة ووحدة المساحة المربعة من باب الملاءمة فقط.

لنفرض أن منطقة مساحتها A يجري مسحها بسحب عينة عشوائية حجمها n من القطع مساحة كل قطعة تساوي a. لنفرض أن $A = Na$ ، كل قطعة تسمى مربع على الرغم من أنها قد لا تكون مربعة المساحة. لنفرض أن m_i تمثل عدد الوحدات في المربع i ونفرض أن M تمثل عدد الوحدات في المجتمع ذو المساحة A حيث إن

$$M = \sum_{i=1}^{N} m_i$$

كذلك نفرض أن

$$\lambda = \frac{M}{A}$$

تمثل كثافة الوحدات أو كثافة الوحدات لكل وحدة مساحة. إن هدفنا هو تقدير λ ومن ثم $M = \lambda A$. لابد من الإشارة هنا إلى أن m_i عبارة عن متغير عشوائي لأنها تمثل عدد الوحدات في مربع ثابت المساحة جرى سحبه بطريقة عشوائية.

لنفرض أننا سحبنا عينة عشوائية حجمها n من المربعات، مساحة كل مربع a وقمنا بمشاهدة عدد الوحدات m_i في كل مربع، لذا فإن الوسط الحسابي

$$\bar{m} = \frac{1}{n}\sum_{i=1}^{n} m_i$$

يُعَدُّ تقديراً لمعدل عدد الوحدات في المربع، وكذلك فإن

$$\hat{\lambda} = \frac{\bar{m}}{a}$$

يمثل تقدير لعدد الوحدات لكل وحدة مساحة، كما نلاحظ أن العينة هنا عبارة عن عينة عشوائية بسيطة؛ لذا يمكننا أن نقدر تباين $\hat{\lambda}$ بسهولة، وهو

$$\hat{V}(\hat{\lambda}) = \frac{1}{a^2}\hat{V}(\bar{m}) = \frac{1}{a^2}\frac{s_m^2}{n}, \quad s_m^2 = \frac{1}{n-1}\sum_{i=1}^{n}(m_i - \bar{m})^2$$

ومن هذه النتيجة نستطيع بسهولة أن نوجد $\hat{M} = \hat{\lambda} A$ ومن ثم نوجد تقديراً لتباين $\hat{M}$.

لقد افترضنا للحصول على النتائج أعلاه أن توزيع الوحدات داخل كل مربع موزعة توزيعاً ثابتاً – أي لا توجد عشوائية في التوزيع– ولكن في كثير

من الحالات الوحدات داخل كل مربع قد تتوزع عشوائياً، على سبيل المثال عدد الأشجار المريضة في وحدة المساحة، عدد البكتريا في وحدة الحجم، عدد الحوادث المرورية في مدة زمنية؛ لذا يمكننا أن نبسط النتائج أعلاه على افتراض أن m_i تتبع توزيع بواسون، لنحصل على تقديرات إلى λ و M.

تقدير للكثافة λ هو

$$\hat{\lambda} = \frac{m}{a}$$

وتقدير تباين $\hat{\lambda}$ يكون

$$\hat{V}(\hat{\lambda}) = \frac{\hat{\lambda}}{an}$$

و تقدير لحجم المجتمع M

$$\hat{M} = \hat{\lambda} A$$

و أخيراً تقدير لتباين $\hat{M}$ هو

$$\hat{V}(\hat{M}) = A^2 \frac{\hat{\lambda}}{an}$$

مثال (4): تعاني إحدى المناطق من انتشار نوع من النمل المؤذي للمزروعات والإنسان، لذا تقدير كثافة انتشار هذا النوع من النمل يعد من أولويات الدوائر الصحية والزراعية في المنطقة. وكذلك لمعرفة الزيادة في حجم مجتمع النمل (أعداد النمل). جري سحب عينة عشوائية تحتوي على 50 مربع، مساحة كل مربع نحو 16 متراً مربعاً، وجرى حصر عدد قرى النمل في كل قطعة وحصلنا على النتائج الآتية:

عدد القرى	0	1	2	3	4	5
التكرار (عدد القطع)	13	8	12	10	5	2

استخدم هذه البيانات لتقدير كثافة قرى النمل في المنطقة.

الحل:

$$\sum_{i=1}^{50} m_i = 0(13) + 1(8) + 2(12) + \ldots + 5(2) = 92$$

و

$$\sum_{i=1}^{50} m_i^2 = 0(13) + 1(8) + 4(12) + \ldots + 25(2) = 276$$

لذا فإن كثافة قرى النمل لكل متر مربع هي

$$\hat{\lambda} = \frac{\bar{m}}{a} = \frac{1}{16}\left(\frac{92}{50}\right) = 0.11$$

ويمكن أن نقدر تباين $\hat{\lambda}$ ليكون

$$\hat{V}(\hat{\lambda}) = \frac{1}{a^2}\frac{s_m^2}{n} = \frac{1}{(16)^2(50)}\frac{1}{49}\left[276 - \frac{(92)^2}{50}\right] = 0.0017$$

وكذلك يمكن إيجاد الحد الأعلى لمقدار الخطأ كما يأتي

$$2\sqrt{\hat{V}(\hat{\lambda})} = 2\sqrt{0.00017} = 0.026$$

وإذا كنا نعتقد أنَّ قرى النمل تتوزع بشكل عشوائي في المنطقة تحت الدراسة، لذا سيكون تقدرينا لتباين $\hat{\lambda}$ بما يلي

$$\hat{V}(\hat{\lambda}) = \frac{\hat{\lambda}}{an} = \frac{0.11}{(16)(50)} = 0.00014$$

و بحد أعلى لمقدار الخطأ هو

$$2\sqrt{\hat{V}(\hat{\lambda})} = 2\sqrt{0.00014} = 0.023$$

ونلاحظ أن تقديري التباين متقاربين جداً وهذا يؤشر إلى أن افتراض عدم العشوائية في توزيع قرى النمل في المنطقة معقول، ولابد من الإشارة هنا إلى أن افتراضنا عدم العشوائية يعني أن $\hat{\lambda}$ و $\hat{V}(\hat{\lambda})$ يعتمدون على البيانات من خلال m فقط.

لمزيد من المعلومات يراجع Scheaffer et al (2005).

تمارين

1. ناقش الفرق بين المعاينة المباشرة والمعكوسة.
2. لنفرض أن تكاليف المعاينة غير مهمة، كيف يمكننا أن نخفض حد الخطأ باستعمال المعاينة المباشرة والمعكوسة.
3. تخشى دائرة الأسماك النهرية من تناقص نوع معين من السمك الذي يعيش في النهر الذي يمر بالولاية التابعة لها، لذا قامت باصطياد r=100 سمكة من هذا النوع من السمك خلال عدة أيام وقامت بوضع علامة مميزة عليها وأعادتها على الفور إلى النهر. بعد عدة أسابيع قامت بسحب عينة ثانية حجمها n=120 فوجدت أن 27 سمكة تحمل العلامة المميزة. قدر عدد الأسماك N الموجودة في النهر من النوع المعني بالدراسة، ومن ثم أوجد حداً أعلى لمقدار الخطأ.
4. ينتاب المسؤولين في إحدى الدوائر قلقٌ من ازدياد عدد الحمام الموجود حول بنايتهم وعلى جدرها الخارجية وسقوفها ومما يسببه من الإزعاج والأوساخ للبناية وساكنيها. لذا قاموا بتأجير فريق يتولى تقدير عدد الحمام الموجود حول البناية. قام الفريق باستخدام عدة طرق لمسك مجموعة من الحمام r = 60 ووضع علامة مميزة على كل حمامة ومن ثم إطلاق سراحها بأسرع وقت ممكن. بعد شهر أعادوا العملية من جديد ولكن مع تحديد أنه لا يمكن إيقاف العملية إلا بعد مسك 18 حمامة تحمل العلامة المميزة. استطاعوا الحصول على العدد المطلوب أي 18 حمامة تحمل العلامة المميزة ولكن بعد مسك 80 حمامة. قدر عدد الحمام الموجود حول البناية ومن ثم أوجد حداً أعلى لمقدار الخطأ.

5. قام الباحثون في إحدى الشركات بسحب عينات بحجم 100 سم^3 من الهواء في المنطقة التي تعمل بها الشركة. الهدف من الدراسة هو معرفة كثافة نوع من البكتريا غير المضرة والموجودة في الهواء. لنفترض أن 15 عينة أعطتنا معدل 210 لكل عينة. قدر كثافة هذا النوع من البكتريا لكل سنتيمتر مكعب ومن ثم أوجد الحد الأعلى للخطأ.
6. ترغب إدارة الشرطة في إحدى المدن بتقدير عدد المتظاهرين الذي خرجوا في مظاهرة منددة بسياسة حاكم المدينة. قامت بتوزيع 80 قميصاً بلون مميز على عينة من المتظاهرين، وطلبت من كل شخص أن يلبس القميص، ومن ثم يعود إلى المظاهرة. بعدة مدة مناسبة قامت إدارة الشرطة بسحب عينة عشوائية ثانية حجمها 60 متظاهراً، فوجدت أنه لا يوجد إلا 5 أشخاص يرتدون القميص ذا اللون المميز في هذه العينة. قدر عدد المتظاهرين، ومن ثم أوجد حداً أعلى لمقدار الخطأ.
7. لنفترض أنه في السؤال رقم (5) من الصعوبة بمكان إيجاد عدد البكتريا في العينة ولكن يمكن بسهولة تمييز فيما إذا كانت البكتريا موجودة أم لا، في عينة حجمها 500 وجدنا أن 0.41 من هذه العينات تحتوي على هذا النوع من البكتريا. قدر الكثافة لكل سنتيمتر مكعب، ومن ثم أوجد الحد الأعلى للخطأ.
8. يجري عد السيارات التي تمر من خلال نقطة تقاطع بين شارعين رئيسين في إحدى المدن خلال مُدَدٍ زمنية محددة، طول كل مدة عشر دقائق، جرى تحديدها عشوائياً خلال ساعات العمل الرسمية، وجرى سحب عينة عشوائية حجمها 20 فترة زمنية، ووجدنا أن معدل عدد السيارات المارة خلال هذه المدة الزمنية ذات العشر دقائق هو 40 سيارة. قدر عدد السيارات المارة خلال ساعات الدوام الرسمي، وأوجد حداً أعلى للخطأ.

References

1. Chapman, D. J. (1952). Inverse, Multiple and Sequential Sample Censuses, *Biometrics*, 8, 286–306.
2. Darling, D. A. and Robbins, H. (1967). Finding the Size of a Finite Population, *Ann. Math. Statist.*, 38, 1392–1398.
3. Scheaffer, R.L., Mendenhall, W. and Ott, L. (2005). Elementary Sampling Survey, 6th ed., Duxbury, New York.
4. Sen, P. K. (1988). Asymptotics in Finite Population Sampling, Handbook of Statistics Vol. 6, pp 291–331, P.R. Krishnaiah and C.R. Rao eds.

الفصل الرابع عشر

المعاينة بطريقة الإمساك وإعادة الإمساك

Capture-Recapture Sampling

1.14 مقدمة

لقد استخدمت طريقة المعاينة بالإمساك وإعادة الإمساك لتقدير أعداد المجتمعات الحيوانية بما فيها الطيور، والأسماك، والحيوانات الثديية، والزواحف، والحشرات وغيرها من المخلوقات. وكذلك لتقدير قابلية اكتشاف الحيوانات في مجتمعاتها لاستخدامها في مسوحات أخرى. وكذلك لتقدير قابلية البقاء ولتقدير معلمات المجتمع الأخرى، كذلك جرى استخدام طريقة الإمساك وإعادة الإمساك لتقدير أعداد المجتمعات الإنسانية المراوغة أو غير المنضبطة مثل أعداد المشردين. وأخيراً تقدير عدد الإصابات المميتة في حوادث الطرق وغيرها.

يمكن تلخيص المعاينة بطريقة الإمساك وإعادة الإمساك كما يأتي: نقوم بسحب عينة عشوائية أولية من أفراد المجتمع، ونضع علامات مميزة على جميع الحيوانات أو نميز هذه الوحدات (الحيوانات) بطريقة أو بأخرى، ونعيدها إلى المجتمع بأسرع وقت، ومن ثم نقوم في وقت لاحق بسحب عينة ثانية مستقلة عن العينة الأولى، ونقوم بحصر عدد الحيوانات التي تحمل العلامة المميزة في هذه العينة، فإذا كانت العينة الثانية ممثلة للمجتمع فإن نسبة الوحدات التي تحمل العلامة المميزة فيها تساوي نسبة الوحدات التي

تحمل العلامة المميزة في المجتمع، وباستخدام هذه العلاقة نستطيع أن نحدد مجموع الوحدات في المجتمع كما سترى لاحقاً.

لابد من الإشارة هنا إلى أنه ليس بالضرورة أن تُمْسَك الحيوانات أو يعاد إمساكها بشكل فعلي، فإذا كان يمكن أن تميز هذه الحيوانات عن طريق بعض العلامات الطبيعية الموجودة عليها فيمكن أن نقوم بالمسح بسحب عينتين مستقلتين، ونستطيع أن نميز عدد الحيوانات المعلمة التي تم مشاهدتها في العينتين، وهذا العدد يمثل إعادة الإمساك، ولكن في كثيرٍ من الحالات لابد من مسك الحيوان باستخدام شبكة أو مصيدة أو أي طريقة يمكن أن يمسك بها، ومن ثم توضع علامات مميزة على الحيوانات مثل قطعة معدنية، أو صبغ، أو شريط، أو علم أو غيرها من العلامات التي نستطيع من خلالها تمييز الحيوانات. أما فيما يخص المجتمعات الإنسانية فالعينتان يمكن أن تكونا عبارة عن كشفين، فمثلاً الكشف الأول يمكن الحصول عليه من التعدادات السكانية الشاملة، والكشف الثاني يمكن أن يكون من دراسة لاحقة لنفس المجتمع، وكذلك يمكن أن يكون الكشف الأول عبارة عن الكشوفات عن الحوادث المرورية لدى مديريات شرطة المرور، والكشف الثاني يمكن أن يكون البيانات المتوافرة لدى شركات التأمين. في كلا المثالين يمكن أن يكون إعادة الإمساك عبارة عن الأشخاص الذين تمت مشاهدتهم في كلا الكشفين.

هناك كثيرٌ من البحوث المنشورة حول هذا النوع من العينات والتي يمكن الرجوع إليها للتوسع في هذا الموضوع، نذكر من هذه البحوث Cormack (1979) و(1978) Otiset al. و(1981, 1991) Pollock و(1990) Pollocket al و(1973, 1982, 1986, 1992) Seber ومن الجدير بالذكر أنَّ Seber (1973) صنف الطرق التي يمكن استخدامها للإمساك وإعادة الإمساك حسب نوع المجتمع. ومن المراجع التي تتناول استخدام طريقة الإمساك وإعادة الإمساك في

المجتمعات الإنسانية المراوغة أو غير المنضبطة Cowan and Malec (1986) و(Freedman (1991 و(Sudman et al. (1988 و(Wolter (1986,1991.

الرموز والمصطلحات الآتية سوف يتم استخدامها في هذا الفصل

τ : عدد الوحدات (الحيوانات) في المجتمع.

X : عدد الوحدات (الحيوانات) التي تحمل العلامات المميزة في المجتمع، وهو كذلك يمثل حجم العينة الأولية.

y : عدد الوحدات (الحيوانات) التي سحبت في العينة الثانية.

x : عدد الوحدات (الحيوانات) التي تحمل العلامات المميزة في العينة الثانية.

وأخيراً لابد من الإشارة إلى عينة بحجم n ترمز إلى مجموعة من الوحدات تم سحبها من المجتمع، وهذه الوحدات إما أن تكون قطع أرض، أو شرائط أرضية، أو خطوطاً مستقيمة، أو مناطق لإمساك الحيوانات أو ... إلخ.

2.14 الإمساك مرة واحدة

يمكن تلخيص طريقة الإمساك وإعادة الإمساك بما يأتي: نقوم بسحب عينة عشوائية أولية من المجتمع بحجم X من الحيوانات، ونقوم بوضع علامات مميزة على جميع هذه الحيوانات ثم نعيدها بأسرع وقت إلى المجتمع، ومن ثم نقوم في وقت لاحق بسحب عينة عشوائية جديدة وبصورة مستقلة عن العينة الأولى وبحجم y من الحيوانات، ونقوم بحصر عدد الحيوانات التي تحمل العلامة المميزة في هذه العينة وليكن x. إذا كانت العينة الثانية تمثل المجتمع بصورة كاملة، فإن نسبة الحيوانات التي تحمل العلامة المميزة في العينة الثانية يكون مساوياً تقريباً لنسبة الحيوانات التي تحمل العلامة في المجتمع أو بعبارة أخرى:

$$\frac{x}{y} = \frac{X}{\tau}$$

وبصورة مماثلة يمكننا القول: إن نسبة الحيوانات التي تحمل العلامة المميزة في المجتمع والتي مسكت في العينة الثانية تساوي تقريباً نسبة العينة الثانية إلى المجتمع، أي:

$$\frac{x}{X} = \frac{y}{\tau}$$

يمكننا حل أي من المعادلتين أعلاه للحصول على تقدير لحجم المجتمع غير المعلوم وهو

$$\hat{\tau} = \frac{y}{x} X$$

هذا التقدير يعرف بتقدير بيترسن، أما تقدير التباين إلى $\hat{\tau}$ فهو

$$s^2(\hat{\tau}) = \frac{Xy(X-x)(y-x)}{x^3}$$

وبفترة 95% ثقة تقريبه لتقدير حجم المجتمع

$$\hat{\tau} \pm 2\sqrt{s^2(\hat{\tau})}$$

بما أن عدد الحيوانات التي تحمل العلامة المميزة في العينة الثانية يمكن أن تكون صفراً، فإنَّه في مثل هذه الحالة لا يمكن تقدير حجم المجتمع؛ لأن تباين $\hat{\tau}$ يكون غير محدد. لذا جرى تعديل تقدير بيترسن $\hat{\tau}$ وتقدير تباينه ليكونا على الوجه الآتي:

$$\tilde{\tau} = \frac{(y+1)(X+1)}{x+1} - 1$$

$$s^2(\tilde{\tau}) = \frac{(X+1)(y+1)(X-x)(y-x)}{(x+1)^2(x+2)}$$

لمزيد من المعلومات يراجع (1982) Seber. وبفترة 95% ثقة تقريبه لتقدير حجم المجتمع

$$\tilde{\tau} \pm 2\sqrt{s^2(\tilde{\tau})}$$

ويمكننا إيجاد فترة ثقة دقيقة وليست بتقريبية وذلك بافتراض التوزيع الاحتمالي، أو باستخدام طريق المحاكاة يراجع (1980, 1984) Buckland. وأيضًا لمزيد من المعلومات حول إيجاد فترة ثقة لتقدير حجم المجتمع يراجع Seber(1973, 1982, 1986, 1992).

لابد من الإشارة هنا إلى أن هذا الجزء من الفصل مشابه إلى حد كبير ما ورد في الفصل الثالث عشر وبالتحديد في 2.13.

مثال (1): قام الباحثون في أحد الحقول بإمساك X=300 فأرة، ووضعت علامة مميزة عليها، ومن ثم أطلق سراحها، وبعد عدة أيام عاد الباحثون إلى حقل الدراسة، وقاموا وبصورة مستقلة عن المرة السابقة بمسك y=200 فأرة، فوجدوا أن من بينها x=50 تحمل العلامة المميزة. استخدم كلا التقديرين أعلاه لتقدير عدد الفئران في الحقل، ومن ثم أوجد 95% فترة ثقة تقريبه لتقدير عدد الفئران في حقل الدراسة.

الحل:

$$\tilde{\tau} = \frac{y}{x}X = \frac{200}{50}(300) = 1,200$$

وبتقدير للتباين

$$s^2(\tilde{\tau}) = \frac{Xy(X-x)(y-x)}{x^3} = \frac{300(200)(300-50)(200-50)}{50^3} = 18,000$$

وبفترة 95% ثقة تقريبه لتقدير حجم المجتمع

$$\hat{\tau} \pm 2\sqrt{s^2(\hat{\tau})} = 1200 \pm 2\sqrt{18000} = (932, 1468)$$

أما إذا استخدمنا التقدير الثاني فنحصل على

$$\tilde{\tau} = \frac{(y+1)(X+1)}{x+1} - 1 = \frac{(200+1)(300+1)}{50+1} - 1 \approx 1185$$

$$s^2(\tilde{\tau}) = \frac{(X+1)(y+1)(X-x)(y-x)}{(x+1)^2(x+2)} = \frac{(300+1)(200+1)(300-50)(200-50)}{(50+1)^2(50+2)} = 16{,}774.5$$

وبفترة 95% ثقة تقريبه لتقدير حجم المجتمع

$$\tilde{\tau} \pm 2\sqrt{s^2(\tilde{\tau})} = 1185.3 \pm 2\sqrt{16774.5} = (926, 1444)$$

3.14 نماذج للإمساك لمرة واحدة

هنالك عدة نماذج رياضية يكمن استخدامها في حالة الإمساك لمرة واحدة سنتطرق إلى أهمها.

أولاً: النموذج المتعدد الحدود Multinomial Model

إن تاريخ مسك أو كشف الحيوان في الصيغة العامة للنموذج المتعدد الحدود لحالة الإمساك أو كشف الحيوان لمرة واحدة في المجتمعات المغلقة يمكن تصنيفه ضمن أربع حالات، هي: الحالة الأولى أن يمسك بالحيوان في العينة الأولية والثانية، الحالة الثانية أن يمسك في العينة الأولى ولا يمسك في الثانية، الحالة الثالثة ألاّ يمسك في العينة الأولى ولكن يمسك في الثانية، وأخيراً الحالة الرابعة ألاّ يمسك الحيوان في العينتين. بعد هذا نرى أن النموذج المتعدد الحدود له أربع احتمالات يكون مجموعها واحداً، ويمكن تطبيقه أو استخدامه لتاريخ إمساك كل حيوان، بالإضافة إلى ذلك إذا كان احتمال

مسك الحيوان في العينتين مستقلاً فإن النموذج لعدد الحيوانات التي لكل واحد منها تاريخ إمساك معين سيكون مضروب النموذج المتعدد الحدود Product Multinomials Model. يمكن أن يكون احتمال كشف أو مسك الحيوان مختلفاً من حيوان لآخر ومن عينة لأخرى، في مثله هذه الحالة يحتوي النموذج العام المتعدد الحدود على عدد كبير من المعالم **Parameters**، ولا يوجد عدد كافٍ من المشاهدات لتقدير هذه المعلمات؛ لذا لابد من وضع بعض القيود حتى نستطيع تقدير حجم المجتمع، ومن أهم هذه القيود أن نفترض أن احتمال كشف أو مسك الحيوان سيكون متساوياً لجميع الحيوانات خلال مدة الدراسة، ولكن يمكن أن يكون الاحتمال مختلفاً بين العينتين، بالإضافة إلى افتراض أن العينتين جرى سحبهما بصور مستقلة. تحت هذه القيود أو الافتراضات نستطيع أن نحصل على تقدير القيمة العظمى لحجم المجتمع $\hat{N} = Xy/x$ Maximum Likelihoodestimator (MLE) وهو نفس تقدير بيترسن الذي مر معنا أعلاه.

ثانياً: نموذج التوزيع الهندسي الزائد Hypergeometric Model

إذا كان عدد الحيوانات التي أُمسكت في العينة الأولى والثانية هو على النحو الآتي X وy ثابتين، والعينة الثانية عبارة عن عينة عشوائية بسيطة دون إرجاع وحدات العينة إلى المجتمع، لذا فإن عدد الوحدات في العينة الثانية التي تحمل العلامة المميزة x سيكون متغيراً عشوائياً ويتبع نموذج التوزيع الهندسي الزائد، مرة أخرى إن تقدير القيمة العظمى لحجم المجتمع سيكون نفس تقدير بيترسن وهو $\hat{N} = Xy/x$.

ثالثاً: نموذج بواسون Poisson Model

يمكن استخدام نموذج بواسون إذا كان عدد الحيوانات في الحالات الأربعة لتاريخ مسك الحيوان هي متغيرات عشوائية مستقلة، وتبعاً لذلك فإن حجم المجتمع τ يكون متغيراً عشوائياً.

لمزيد من المعلومات حول نماذج الإمساك وإعادة الإمساك يراجع Cormack (1979) و Otis et al. (1978) و Pollock (1981) و Seber (1982, 1986,1992) و Pollock et al. (1990) و Wolter (1986).

4.14 تصميم المعاينة للإمساك وإعادة الإمساك

لقد افترضنا في النماذج أعلاه أن مسك أو كشف الحيوان يكون مستقلاً عن مسك الحيوانات الأخرى، أو أن العينة الثانية هي عبارة عن عينة عشوائية بسيطة مع عدم الإرجاع كما هو الحال مع النموذج الهندسي الزائد. عندما يتم الحصول على المشاهدات باستخدام عينة من الوحدات مثل قطع أراض، أو مواضع مصائد للحيوانات، أو شبكات لمسك الأسماك (الطيور) أو خطوط عرضية أو ... إلخ لا تكون العينة عينة عشوائية بسيطة، ولا تتم عملية كشف أو مسك الحيوانات مستقلة كما افترضنا في النماذج أعلاه. لمثل هذه الحالات (النماذج السابقة) تكون مفيدة لمعرفة التغيرات التي لا تعتمد على العينة، ولكن تغيرات المعاينة التي تعتمد على الطريقة التي حصلنا بها على العينة تتأثر ببعض العوامل مثل عدم تجانس المكان، وكذلك الفروقات بين المجموعات، والتي من الممكن أن تكون مسؤولة عن جزء كبير من تباين تقدير بيترسن $\hat{\tau}$.

من الناحية المثالية نرغب بأن تكون الحيوانات في منطقة الدراسة مثل حبات الفاصوليا الموجودة في الصحن. لتقدير عدد حبات الفاصوليا في

الصحن، نقوم بأخذ مجموعة من هذه الحبات ونضع علامة على كل حبة باستخدام القلم أو نلونها بلون مميز، ومن ثم نعيدها إلى الصحن ونخلطها مع الحبات الأخرى حتى تكون الحبات متجانسة داخل الصحن. ومن ثم نقوم بسحب عينة ثانية وهي التي يمكن اعتبارها عينة عشوائية، أي كل حبه لديها نفس الاحتمال بالظهور في العينة الثانية، وكل مجموعة حبات (عينة) لها نفس الاحتمال بالسحب. إذا كان المجتمع الحيواني الذي نرغب دراسته يمتاز بنفس الصفة أعلاه أي نستطيع أن نسحب عينة عشوائية بسيطة منه، فإن النموذج الهندسي الزائد يمكن استخدامه في مثل هذه الحالة، ويكون المتغير العشوائي هو عدد الحيوانات المعلمة (أي تحمل العلامة المميزة) في العينة الثانية. يمكن أن نستخدم النموذج المتعدد الحدود في مسوحات الحياة البرية التي يتم سحب العينة فيها بالطيران فوق منطقة الدراسة باستخدام مروحية، التي تؤدي إلى وضع علامة مميزة على جميع الحيوانات التي تم اكتشافها في العينة الأولى ولا يتم تعليم أي حيوان في العينة الثانية، لذلك نرى أن كلا العينتين يكون حجمها غير ثابت. في مثل هذه المسوحات يمكن أن نفترض أن سحب أي حيوان يكون مستقلاً عن سحب الحيوانات الأخرى، لهذا فإن استخدام نموذج مضروب متعدد الحدود يكون مناسباً.

الآن لنتصور أن عملية الإمساك وإعادة –الإمساك في المجتمعات الحيوانية– التي يتم الحصول فيها على العينة الأولية بواسطة مصائد وضعت في أماكن مختلفة من منطقة الدراسة. ومع كل مصيدة هنالك مناطق بمساحات مختلفة لجلب انتباه الحيوان لدخولها ومن ثم مسكه، وكذاك حركة الرياح وحركة الحيوانات كلها تؤثر على عملية مسك الحيوان، ولكن الحيوانات المعادة إلى المجتمع بعد مسكها ووضع العلامة المميزة عليها يمكن أن تتوزع في المجتمع بصورة غير متجانسة، وربما له علاقة بالمنطقة التي مسكت بها أو المنطقة التي جرى إطلاق سراحها فيها، والذي قد يؤدي

إلى أن يكون توزيع الحيوانات المعلمة مخالفاً لتوزيع الحيوانات في المجتمع. ونتيجة لذلك سيكون نسبة الحيوانات المعلمة في منطقة معينة يختلف كثيراً عن المناطق الأخرى.

حسب النماذج المعتمدة التي تكلمنا عنها أعلاه لابد أن تكون العينة الثانية عبارة عن عينة عشوائية بسيطة من الحيوانات، أو أن عملية مسك أو كشف الحيوان يكون مستقلاً عن الحيوانات الأخرى، ولكن في كثير من الحالات يكون هذا غير ممكن. بدلاً من ذلك يمكن أن نحصل على عينة ثانية من الحيوانات من خلال عينة من وحدات تسمى وحدات الكشف وتكون عبارة عن وحدات من الأرض، أو مصائد، أو قطع أرض، أو شبكات صيد أو غيرها من وحدات الكشف، في مثل هذه الحالات تكون الوحدات موزعة بشكل عشوائي على المجتمع، ولكن الحيوانات التي تم مسكها لا تكون مستقلة عن بعضها الآخر، وبما أن الحيوانات المعلمة غير موزعة بصورة متساوية في المجتمع؛ لذا سنرى أن بعض الوحدات ستحتوي على نسبة كبيرة من الحيوانات المعلمة بينما وحدات أخرى تكون نسبة الحيوانات المعلمة قليلة جداً أو صفراً.

فيما يأتي سنقوم باستخدام عينة عشوائية من وحدات الكشف والتي يتم سحبها مع الإرجاع أو دون إرجاع مع عملية الإمساك وإعادة الإمساك لتقدير حجم المجتمع.

أولاً: العينة العشوائية مع الإرجاع لوحدات الكشف

لنفترض أننا سحبنا عينة عشوائية (مع الإرجاع) حجمها n من وحدات الكشف على سبيل المثال قطع أراض أو مصائد أو شبكات صيد أو ... إلخ. لنفترض أن للوحدة i من وحدات الكشف y_i والتي تمثل عدد الحيوانات التي شوهدت في الوحدة و x_i عدد الحيوانات المعلمة (أي تحمل العلامة المميزة) من

بين y_i ؛ لذا فإن عدد الوحدات التي تم مشاهدتها في العينة الثانية هو $\sum_{i=1}^{n} y_i$ من بينهم $\sum_{i=1}^{n} x_i$ حيوان يحمل العلامة المميزة. لنفترض أن $\bar{y}$ و $\bar{x}$ تمثلان الوسط الحسابي للمتغيرين y_i و x_i على التوالي. وكما نعلم فإن عدد الوحدات المعلمة في المجتمع X معلومة، إذا افترضنا أن المجتمع مغلق، أي عدد الحيوانات المعلمة بقي كما هو منذ أن تم تنفيذ العينة الأولى إلى أن تنفذ العينة الثانية؛ لذا فإن تقدير حجم المجتمع يكون

$$\hat{\tau} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} X = \frac{\bar{y}}{\bar{x}} X$$

وهو نفس تقدير النسبة Ratio المعروف، أما تقدير التباين إلى $\hat{\tau}$ فإن نفس تقدير تباين النسبة المعدل والمعروف أيضاً

$$s^2(\hat{\tau}) = \left(\frac{X}{\bar{x}}\right)^2 \frac{1}{n(n-1)} \sum_{i=1}^{n} (y_i - r x_i)^2$$

حيث إن $r = \frac{\bar{y}}{\bar{x}}$. وأخيراً فترة $100(1-\alpha)\%$ ثقة تقريبية لتقدير حجم المجتمع يمكن إيجادها باستخدام

$$\hat{\tau} \pm t_{\alpha/2} \sqrt{s^2(\hat{\tau})}$$

حيث إن $t_{\alpha/2}$ يمكن إيجادها من جداول توزيع t مع درجة حرية $n-1$. لمزيد من المعلمات ولصيغ أخرى لتقدير التباين يراجع Seber (1982).

سيكون تباين تقدير النسبة صغيراً عندما تتوزع الحيوانات المعلمة داخل منطقة الدراسة بصورة متساوية، وعدم توزيع الحيوانات المعلمة بصورة متساوية سيؤدي إلى زيادة التباين. أخيراً إذا كان توزيع الحيوانات غير متساوٍ لدرجة كبيرة قد يؤدي هذا إلى جَعْل تباين $\hat{\tau}$ غير محدد.

مثال (2): قام الباحثون بمسك 382 طير في مجتمع للطيور بواسطة شباك شفافة، ومن ثم قاموا بوضع علامة براقة على جميع الطيور التي مُسِكت، يستطيع المشاهد أن يرى هذه العلامات بواسطة منظار من مسافة بعيدة. العينة الثانية تم تنفيذها باستخدام ثماني خطوط محددة الأطوال، جرى وضعها بطريقة عشوائية في مجتمع الدراسة. يقوم المشاهد أو الباحث بالتحرك على كل خط ويحصي عدد الطيور التي يشاهدها بواسطة المنظار في أثناء سيره على الخط ويحصي من بينها عدد الطيور التي تحمل الإشارة المميزة. الجدول الآتي يبين عدد الطيور التي تم مشاهدتها **y** وعدد الطيور المعلمة **x** لكل خط

الخط	1	2	3	4	5	6	7	8
Y	33	13	1	45	21	82	14	0
X	20	2	0	15	5	39	4	0

استخدم تقدير النسبة لتقدير حجم المجتمع ومن ثم أوجد 95% فترة ثقة لتقدير حجم المجتمع.

الحل: نسبة العينة هي

$$r=\frac{\bar{y}}{\bar{x}}=\frac{26.125}{10.625}=2.459$$

تقدير عدد الطيور

$$\hat{\tau}=\frac{\bar{y}}{\bar{x}}X=rX=2.459(382)\approx 939$$

تقدير التباين

$$\sum_{i=1}^{n}(y_i-rx_i)^2=[33-2.459(20)]^2+\ldots+[0-0(2.459)]^2=680.318$$

$$s^2(\hat{\tau})=\left(\frac{X}{\bar{x}}\right)^2\frac{1}{n(n-1)}\sum_{i=1}^{n}(y_i-rx_i)^2=\left(\frac{382}{10.625}\right)^2\frac{1}{8(7)}(680.318)=15,703.37$$

أما 95% فترة ثقة تقريبية لتقدير حجم المجتمع فتكون

$$\hat{\tau}\pm t_{\alpha/2}\sqrt{s^2(\hat{\tau})}=939\pm 2.365\sqrt{15,703.37}=(643,1235)$$

ولابد من الإشارة إلى أن تقدير التباين لتقدير بيترسن سيكون

$$s^2(\hat{\tau}) = \frac{Xy(X-x)(y-x)}{x^3} = \frac{382(209)(382-85)(209-85)}{85^3} = 4,789$$

وهو أقل بكثير من تقدير قيمة التباين التي حصلنا عليها هنا، والسبب يعود إلى أن تقدير بيترسن لم يأخذ في الحسبان عدم التجانس المكاني داخل المجتمع، وكذلك لا توجد استقلالية بين الطيور كما يفترضها التقدير. على هذا يكون التقدير هنا أكثر دقة لتقدير التباين الحقيقي لتقدير حجم المجتمع.

ثانياً: العينة العشوائية مع عدم الإرجاع لوحدات الكشف

يمكن استخدام العينة العشوائية من دون إرجاع الوحدات وذلك عندما تكون الوحدات في الكشف عبارة عن مجموعات غير متداخلة أو عندما تكون منطقة الدراسة مقسمة إلى مجموعة من المناطق الصغيرة، كل منطقة تمثل وحدة مستقلة مع كون احتمالات كشف جميع الحيوانات متساوية في جميع الوحدات التي تم سحبها في العينة. إن طريقة الإمساك وإعادة الإمساك تتضمن سحب عينة ثانية بحجم n من وحدات الكشف التي يبلغ عددها بالمجتمع N من دون إرجاع مع احتمال كشف أو مسك العناصر (الحيوانات) ثابتة لجميع العناصر الموجودة في وحدة الكشف كما هو موصوف من قبل Seber (1982, pp.111-114) و Wolter (1986). إن تقدير حجم المجتمع τ سيكون نفس تقدير بيترسن وهو

$$\breve{\tau} = \frac{y}{x} X$$

و مع تقدير للتباين كما اقترح من قبل Wolter (1986) وهو

$$s^2(\breve{\tau}) = \left(\frac{X}{\bar{x}}\right)^2 \frac{N-n}{Nn(n-1)} \sum_{i=1}^{n} (y_i - r x_i)^2 + \frac{X \bar{y}(m_x - \bar{x})(\bar{y} - \bar{x})}{\bar{x}^3}$$

حيث إن $\mu_x = \frac{X}{N}$. يراجع Wolter (1986) لمزيد من المعلومات.

5.14 تقدير قابلية الاكتشاف باستخدام نماذج الإمساك وإعادة الإمساك

إن استخدام طريقة الإمساك وإعادة الإمساك لتقدير حجم المجتمع تقوم بصورة مباشرة أو غير مباشرة بتقدير احتمالات الإمساك أو الاكتشاف، على سبيل المثال فإن تقدير بيترسن يشترك مع تقديرات أخرى تستعمل لتقدير حجم المجتمع الحيواني بالعلاقة نفسها التي يمكن كتابتها بالشكل الآتي

$$\hat{\tau}=\frac{y}{\hat{p}}$$

حيث إن $\hat{p}=\frac{x}{X}$ هو عبارة عن تقدير احتمال الاكتشاف أو الإمساك في العينة الثانية، كذلك يمكن استخدام طريقة الإمساك وإعادة الإمساك لتقدير الاكتشاف لمسوحات أخرى مستقلة يمكن تنفيذها بنفس طريقة المعاينة المتبعة في العينة الثانية.

إذا افترضنا التوزيع الهندسي الزائد إلى x، فإن التباين لتقدير الاكتشاف إذا عرفنا X وy سيكون

$$\text{var}(\hat{p})=\frac{y}{X^2}\left(\frac{X}{\tau}\right)\left(1-\frac{X}{\tau}\right)\frac{\tau-y}{\tau-1}$$

إذا تم اختيار n من وحدات الكشف (مع الإرجاع) مثل قطع أراض، أو خطوط مستقيمة، أو مصائد أو شبكات صيد ... إلخ، ولنفترض أن x_i تمثل عدد الحيوانات التي تحمل العلامة المميزة في الوحدة i، لذا فإن تقدير احتمال الاكتشاف لكل وحدة سيكون

$$\hat{p}=\frac{x'}{X}$$

حيث إن $x'=\sum_{i=1}^{n}x_i$. أما تقدير التباين إلى $\hat{p}$ فيكون

$$s^2(\hat{p})=\frac{s_x^2}{nX^2}$$

حيث إن

$$s_x^2 = \sum_{i=1}^{n}(x_i - \bar{x})^2 /(n-1)$$

مثال (3): بالعودة إلى المثال 1 قدِّر احتمال الإمساك أو الاكتشاف للفئران في العينة الثانية، ومن ثم أوجد تقدير الانحراف المعياري لهذا التقدير.

الحل: احتمال إعادة الإمساك أو الاكتشاف هو

$$\hat{p} = \frac{x}{X} = \frac{50}{300} = 0.17$$

أما تقدير التباين لهذا التقدير فيمكن إيجاده باستخدام $\hat{\tau}$ بدلاً من τ لنحصل على

$$s^2(\hat{p}) = \frac{y}{X^2}\left(\frac{X}{\hat{\tau}}\right)\left(1 - \frac{X}{\hat{\tau}}\right)\frac{\hat{\tau} - y}{\hat{\tau} - 1} = \frac{200}{300^2}\left(\frac{300}{1200}\right)\left(1 - \frac{300}{1200}\right)\frac{1200 - 200}{1200 - 1} = 0.00035$$

وهذا يعطينا تقديراً للانحراف المعياري مقداره 0.02.

قام عدد من الباحثين مثل (1990) Ahlo و(1983) Pollock and Otto و(1969) Burnham and Oberton و(1969) Overton باقتراح تقديرات يمكن استخدامها لتقدير حجم المجتمع في الشكل الآتي

$$\hat{\tau} = \sum_{i=1}^{\tau} z_i / \hat{\pi}_i$$

حيث إن $z_i = 1$ إذا كانت الوحدة i في المجتمع جرى سحبها في عينة أو أكثر و $z_i = 0$ خلاف ذلك و $\hat{\pi}_i$ تمثل تقدير احتمال اكتشاف أو مسك الحيوان في العينة الأخيرة. لقد تم تقدير الاحتمالات $\hat{\pi}_i$ بطريق مختلفة منها مباشرة وأخرى غير مباشرة، وهذا قد يؤدي إلى تحيز كبير بتقدير الاحتمالات، لمزيد من المعلومات يراجع (1978) Otis et al و Pollock and

Otto(1983). كذلك يراجع (1990) .Pollock et al لمعرفة كيفية تقدير حجم المجتمع إذا كان احتمال كشف أو مسك الحيوانات مختلف.

6.14 نماذج أخرى

باستخدام عينات متتابعة من مجتمع يتعرض أفراده للموت،أو الولادة لعناصر جديدة، أو الهجرة خارج المجتمع، أو العودة إلى المجتمع، هذا يعني أن المجتمع لن يكون مغلقا ومن ثمَّ لابد من الأخذ في الحسبان هذه التغيرات التي قد تحدث خلال مدة سحب العينات المتتابعة، كل ذلك يتطلب نماذج قد تحتوي على عدد كبير من المعالم. أحد النماذج الذي يسمح بموت بعض أعضاء المجتمع والذي سببه عملية الإمساك وإعادة الإمساك اُقترح من قبل Jolly (1965) وSeber(1965) ولقد تم توسيع هذا النموذج بطرق مختلفة من قبل عدد كبير من الباحثين، يراجع (Seber(1986. في هذا النموذج الاحتمالات يمكن أن تتغير من عينة إلى أخرى ولكنها تبقى ثابتة لجميع الحيوانات لكل مسك، لقد تم استخدام النموذج الهندسي الزائد المتعدد المتغيرات. لقد قام Seber(1982,1986,1992) بمراجعة جميع البحوث المنشورة حول هذا النموذج وأضاف تعديلاته المختلفة.

هناك طريقة جديدة نوعاً ما لتحليل طريق الإمساك وإعادة الإمساك للمجتمعات المفتوحة باستخدام نماذج log‑linear، لمزيد من التفاصيل يراجع(Cormack (1980, 1981 , 1985,1989 وAgresti(1990). في هذه النماذج لوغاريتم توقع عدد الحيوانات مع تاريخ إمساك يكون عبارة عن دالة خطية لمجموعة من المعالم التي يمكن تفسيرها.

References

1. Agresti, A. (1990). Categorical Data Analysis, New York, Wiley.
2. Ahlo, J. M. (1990). Logistic Regression in Capture-recapture Models. *J. Amer. Statist. Assoc.* 46, 623-635.
3. Buckland, S. T. (1980). A Modify Analysis of the Jolly-Seber Capture-recapture Model. *Biometrics*, 36, 419-435.
4. Buckland, S. T. (1984). Monte Carlo Confidence Intervals. *Biometrics*, 40, 811-817.
5. Burnham, K. P. and Overton, W. S. (1978). Estimation of the Size of a Closed Population when Capture Probability Vary Among Animals. *Biometrika*, 65, 625-633.
6. Cormack, R. M. (1979). Models for Capture-recapture. In R. M. Cormack, G. P. Patil, and D. S. Robson (eds.), *Sampling Biological Populations*. Fairland, MD: International Co-operative Publishing House, 217-255.
7. Cormack, R. M. (1980). Model Selection in Capture-recapture Experiments. *GLIM Newsletter, Dec.*, 27-29.
8. Cormack, R. M. (1981). Loglinear Models for Capture-recapture Experiments on Open Populations. In R.W. Hiorns and D. Cooke (eds.). *The Mathematical Theory of the Dynamics of Biological Populations II*. London, Academic Press, 243-273.
9. Cormack, R. M. (1985). Examples of the use of GLIM to Analyse Model Capture-recapture Studies. In B.J.T. Morgan and P. M. North (eds.) *Statistics in Ornithology*> Lecture Notes in Statistics, No. 29. New York, Springer-Verlag, 243-273.
10. Cormack, R. M. (1989). Log-linear Models for Capture-recapture. *Biometrics*, 45, 395-413.
11. Cowan, C. D. and Malec, D. (1986). Capture-recapture Models when both Sources have Clustered Observations. *J. Amer. Statist. Assoc.* 81, 347-353.
12. Freedman, D. A. (1991). Adjusting the 1990 Census. *Science*, 252, 1233-1236.
13. Jolly, G. M. (1965). Explicit Estimates from Capture-recapture with both Death and Immigration-stochastic Model. *Biometrika*, 52, 225-247.
14. Otis, D. L., Burnham, K. P., White, G. C. and Anderson, D.R. (1978). Statistical for Capture Data from Closed Population. *Wildlife Monographs*, No. 2.
15. Otto, M. C. and Pollock, K. H. (1990). Size Bias in Transect Sampling: A filed Test. *Biometrics*, 239-245.
16. Overton, W. S. (1969). Estimating the Number of Animals in Wildlife Population. In R. H. Giles (ed). *Wildlife Management Techniques*, 3rd ed. Washington, DC, Wildlife Society, 403-455.
17. Pollock, K. H. (1981). Capture-recapture Models: A Review of Current Methods, Assumptions and Experimental Design. In C. J. Ralph and J. M. Scott (eds.), *Estimating Numbers of Terrestrial Birds*, Studies in Avian Biology, No. 6 Oxford, Pergamon Press, 426-435.
18. Pollock, K. H. (1991). Modeling Capture-recapture, and Removal Statistics for Estimation of Demographic Parameters for Fish and Wildlife Populations: Past Present and Future. *J. Amer. Statist. Assoc.* 86, 225-238.
19. Pollock, K. H., Nichols, J. D., Hines, J. E. and Brownie, C. (1990). Statistical Inference for Capture-recapture Experiments, *Wildlife Monograph*, 107, 1-97.

20. Pollock, K. H. and Otto, M. C. (1983). Robust Estimation of a Population Size in Closed Animal Population from Capture–recapture Experiments. *Biometrics*, 39, 1035–1049.
21. Seber, G. A. F. (1965). A Note on the Multiple–recapture Census. *Biometrika*, 49, 339–349.
22. Seber, G. A. F. (1973). The Estimation of the Animal Abundance, London, Griffin.
23. Seber, G. A. F. (1982). The Estimation of the Animal Abundance, 2nd ed. London, Griffin.
24. Seber, G. A. F. (1986). A Review of Estimating Animal Abundance. *Biometrics*, 42, 267–292.
25. Seber, G. A. F. (1992). A Review of Estimating Animal Abundance. *II International Statistical Review*, 60, 129–166.
26. Sekar, C. C. and Deming, W. E. (1949). On a Method of Estimating Birth and Death Rates and the Extent of Regression. *J. Amer. Statist. Assoc.* 5, 119–127.
27. Sudman, S., Sirken, M. G. and Cowan, C. D. (1988). Sampling Rare and Elusive Populations. *Science*, 240, 991–996.
28. Thompson, S. K (2002). Sampling, 2end Wiley, New York.
29. Wolter, K. M. (1986). Some Coverage Error Models for Census Data. *J. Amer. Statist. Assoc.* 81, 338–346.
30. Wolter, K. M. (1991). Accounting for America's Uncounted and Miscounted. *Science*, 253, 12–15.

الفصل الخامس عشر

المعاينة بالخط القاطع

Line Transect Sampling

1.15 مقدمة

إن المسح بطريقة الخط القاطع للحيوانات أو النباتات الموجودة في منطقة الدراسة، يجرى عادة بأن يقوم العَاد أو الباحث بالتحرك على الخط الذي يجري اختياره ليقطع منطقة الدراسة، ويقوم بمشاهدة الحيوانات أو النباتات على جانبي الخط، ويحدد موقعها، وبعد ذلك يقوم بقياس المسافة العمودية بين موضع الحيوان أو النبات والخط القاطع، عادة ما تتم مشاهدة أعداد كبيرة من الحيوانات أو النباتات القريبة من الخط ليس لأن كثافتها قرب الخط أكبر من المناطق البعيدة، ولكن احتمال اكتشاف أو مشاهدة الحيوانات أو النباتات القريبة من الخط تكون أكبر، ويتناقص بشكل تدريجي كلما ابتعدنا عن الخط القاطع، ولابد من الأخذ في الحسبان حالة عدم ثبات اكتشاف الحيوانات أو النباتات عندما نقوم بتقدير عدد أو كثافة الحيوانات في منطقة الدراسة.

لقد جرى استخدام طريقة الخط القاطع لمجتمعات متعددة من ضمنها مجتمعات الطيور والحيوانات وأصناف كثيرة من النباتات، كذلك للمجتمعات التي تكون عملية اكتشاف أفرادها يعتمد على موضع أو مكان العَاد أو الباحث، وكذلك جرى استخدام الخط القاطع للمسوحات الجوية باستخدام الطائرات، والمسوحات البحرية باستخدام القوارب، والمسوحات

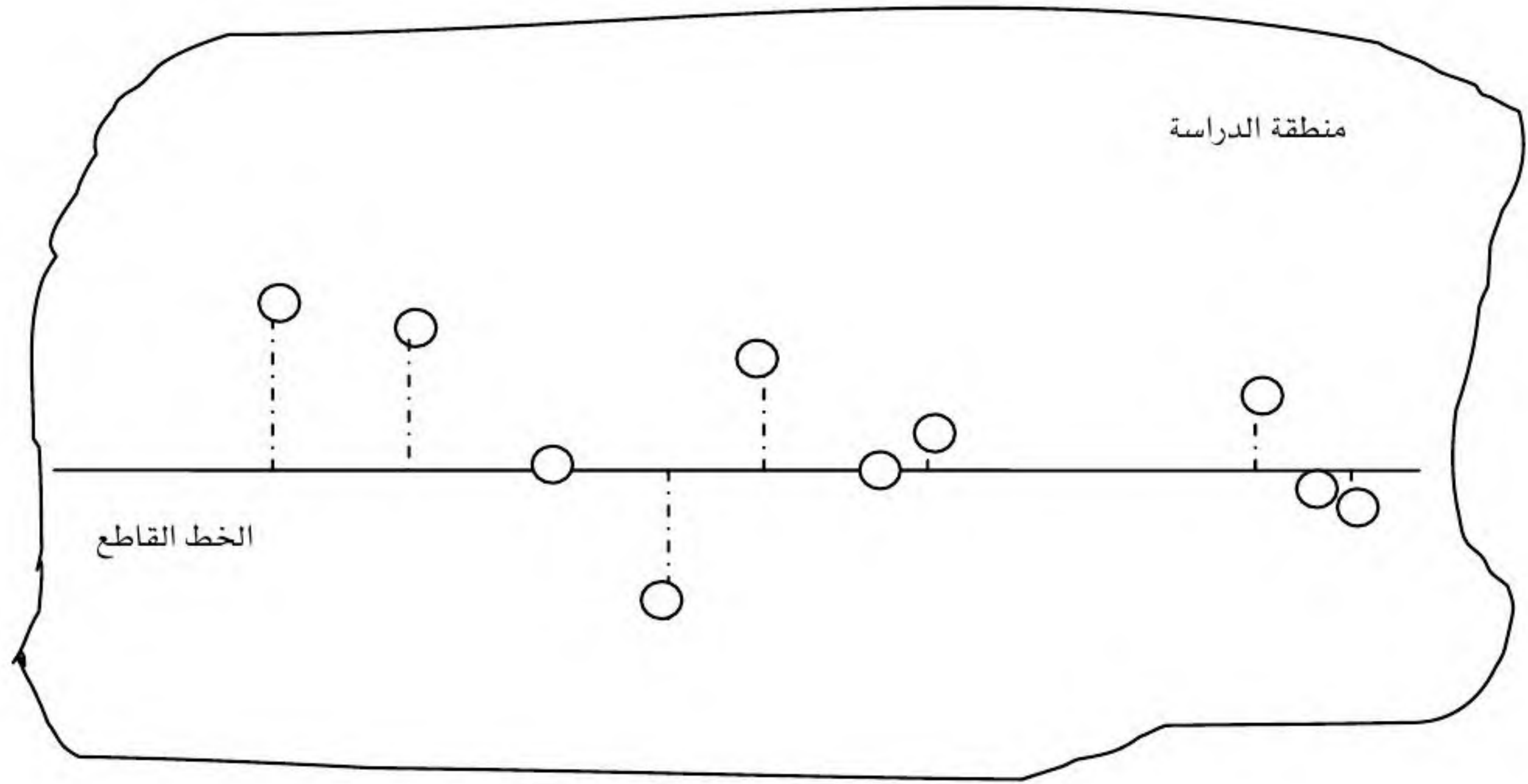
البرية باستخدام السيارات. وللتسهيل سوف نستخدم عبارة "حيوان أو حيوانات" للتعبير عن وحدات المجتمع الذي يجري دارسته، وقد تكون هذه الوحدات غير حيوانات.

عادة ما يوصف المسح بطريقة الخط أو الخطوط القاطعة بدالة الاكتشاف التي تعطي احتمال أن تتم عملية اكتشاف الحيوان (النبات) وهو في موضع معين من الخط القاطع، وعادة ما يتناقص احتمال اكتشاف الحيوان كلما كان موقعه من الخط أبعد، ويكون احتمال اكتشاف الحيوان (النبات) مساوياً لواحد إذا كان على الخط في أغلب الأحيان.

لمزيد من المعلومات ولتفاصيل أكثر يراجع: Gates (1979) و Buckland et al. (1992) و Burnham et al. (1980) و Eberhardt(1978) و Seber (1982, 1986, 1992) و Ramsey al al. (1988).

2.1.5 طرق تقدير الكثافة

الشكل 1: يوضح الحيوانات أو أي عناصر أخرى جرى مشاهدتها من على الخط القاطع، المسافة العمودية من العناصر إلى الخط القاطع جرى توضيحها بالخطوط المتقطعة. السؤال المهم هو كيف يمكننا أن نقدر كثافة العناصر في منطقة الدراسة إذا كانت لدينا المسافات العمودية من العناصر إلى الخط أو الخطوط القاطعة وكانت قابلية الاكتشاف للعناصر على الخط أو قربه كاملة، أي نستطيع أن نكتشف جميع الحيوانات أو العناصر، ولكن قابلية الاكتشاف تتناقص كلما ابتعدنا من الخط. سوف نبدأ باستخدام هذه المعلومات لتقدير كثافة الحيوانات أو العناصر وذلك باستخدام بعض الطرق البسيطة أو البدائية، ومن ثم ننتقل إلى الطرق الأكثر تقدماً.



الشكل [1] مشاهدة الحيوانات (العناصر) من الخط القاطع

إن الهدف الرئيس من الطرق التي سنتكلم عنها هنا هو تقدير كثافة الحيوانات (العناصر) في منطقة الدراسة. فيما يأتي بعض الرموز التي سنحتاجها في هذا الفصل:

A : مساحة منطقة الدراسة.

y_i : عدد الحيوانات (العناصر) التي تم مشاهدتها من الخط القاطع i.

n : حجم العينة ويكون هنا عدد الخطوط القاطعة التي جرى اختيارها في منطقة الدراسة.

T : عدد الحيوانات (العناصر) في منطقة الدراسة.

$D=T/A$: كثافة الحيوانات (العناصر) في منطقة الدراسة.

من الجدير بالذكر هنا أن Burnham et al. (1980) اقترح أن يكون عدد الحيوانات (العناصر) التي جرى مشاهدتها لا يقل عن 40 من أجل الحصول

على تقدير يمكن الاعتماد عليه لكثافة الحيوانات أو أي عناصر أخرى في منطقة الدراسة.

1.215 طريقة الشريط الضيق (Narrow Strip Method)

على الرغم من كون قابلية اكتشاف الحيوانات البعيدة من الخط تكون غير كاملة، فإنه لابد من وجود شريط ولو ضيق حول الخط تكون فيه قابلية الاكتشاف كاملة. باستخدام العناصر الموجودة داخل الشريط وبإهمال العناصر الموجدة خارجه يمكننا أن نقدر عدد عناصر المجتمع أو الكثافة لهذه العناصر في المجتمع باستخدام الطريقة الآتية:

لنفترض أن L تمثل طول الخط و w_0 تمثل المسافة القصوى من الخط والتي تكون قابلية الاكتشاف فيها كاملة. إذًا سيكون عرض الشريط $2w_0$ ومساحته $2w_0L$. لنفترض أن y_0 يمثل عدد الحيوانات أو العناصر التي تم مشاهدتها في الشريط. لتقدير الكثافة D في الشريط الضيق يمكننا أن نستخدم

$$\hat{D}=\frac{y_0}{2w_0L}$$

إذا كانت مساحة منطقة الدراسة A، فإن عدد الحيوانات في المجتمع يمكن تقديره باستخدام

$$\hat{T}=A\hat{D}=\frac{Ay_0}{2w_0L}$$

بصورة عامة تكون المسافة w_0 أقل من المسافة القصوى التي يمكن أن تكتشف فيها العناصر، لذلك فإن y_0 يكون أقل من عدد العناصر المكتشفة، هنالك طرق عديدة جرى إتباعها لاختيار المسافة w_0 يكون فيها

الاكتشاف كاملاً، أحد هذه الطرق تكون بدراسة المدرج التكراري لمسافة الاكتشاف، وتجري ملاحظة المسافة التي ينحدر عندها التكرار النسبي للمشاهدات بصورة حادة.

مثال 1: لتقدير كثافة الطيور في إحدى المناطق قام الباحث باستخدام الخط القاطع وبطول $L=100$ متر، وتم اكتشاف $y=18$ طير بالمسافات الآتية من الخط:

$$0, 0, 1, 3, 7, 11, 11, 12, 15, 15, 18, 19, 21, 23, 28, 33, 34, 44$$

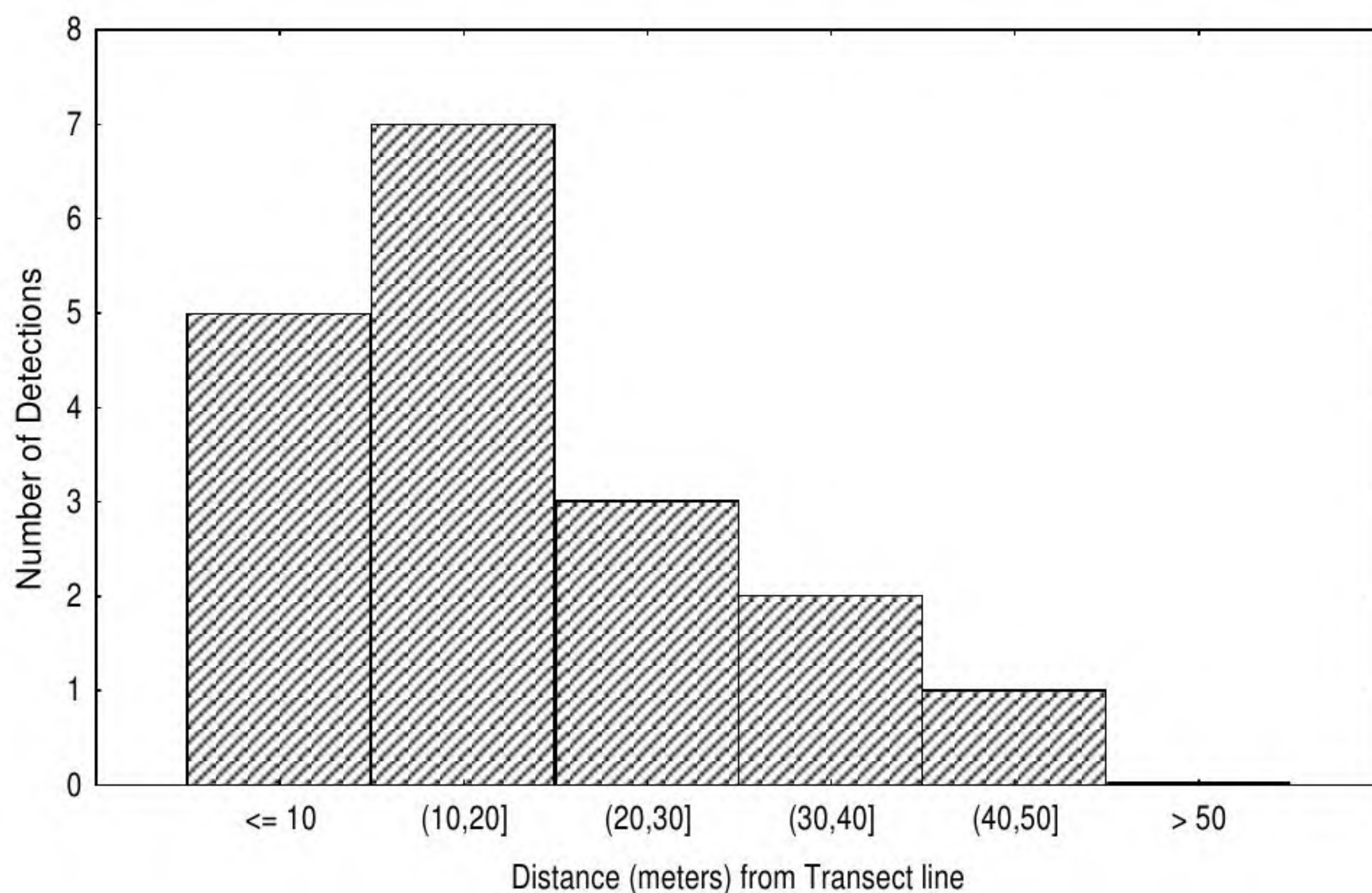
الحل: إذا قمنا برسم المدرج التكراري شكل2 بفئات أطوالها 10 متر، نلاحظ أن 5 طيور تم مشاهدتها في الفئة الأولى، و7 طيور في الفئة الثانية ما بين 10 إلى 20 متر، و3 طيور في الفئة من 20 إلى 30 متر، وهكذا، نلاحظ أن عدد الطيور المشاهدة بعد الفئة الثانية قد انخفض بشكل ملحوظ؛ لذا اخترنا $w_0 = 20$، وعليه سيكون عرض الشريط الضيق $2w_0$ يساوي 40م. لذا فإن عدد الطيور التي يمكن مشاهدتها في الشريط الضيق هي $y_0 = 12$. وباستخدام طريقة الشريط الضيق يكون تقدير كثافة الطيور

$$\hat{D} = \frac{y_0}{2w_0L} = \frac{12}{2(20)(100)} = 0.003$$

لذا فإن تقدير كثافة الطيور بالمتر المربع هي 0.003 طير.

على الرغم من أن طريقة الشريط الضيق سهلة جداً، إلا أنها غير مرضية لعدة أسباب منها:

1. لم يتم استخدام جميع المشاهدات التي جرى جمعها.
2. إن تحديد عرض الشريط الضيق جرى بشكل اعتباطي.
3. وأخيراً لأن قابلية اكتشاف الحيوانات لن تتناقص بشكل منتظم مع المسافة، لذا فإن عرض الشريط الضيق مع قابلية اكتشاف كاملة قد يكون صفراً.



الشكل2: المدرج التكراري لعدد الطيور التي تم اكتشافها لكل10م من الخط القاطع

2.2.1.5 الطرق المعلماتية (Parametric Methods)

عندما تتم مشاهدة الحيوانات في شريط من الأرض، هذا يعني أن جميع الحيوانات في الشريط الذي نصف عرضه يساوي w وطوله يساوي L قد تمت مشاهدتها، لذا فإن تقدير الكثافة يكون $\hat{D} = y/2Lw$ أي عدد الحيوانات مقسوماً على مساحة الشريط، عندما تتم مشاهدة الحيوانات من الخط القاطع بدالة اكتشاف g(x) لديها قابلية اكتشاف كاملة على الخط وتتناقص بالبعد عن الخط بمسافة x، فإن المسافة لمشاهدة الحيوانات من الخطوط التي جرى وضعها في منطقة الدراسة بشكل عشوائي سيكون لديها دالة كثافة الاحتمال f(X) وبنفس الشكل لدالة الاكتشاف g(x) ولكن

المساحة تحت دالة كثافة الاحتمال $f(X)$ تساوي واحداً، وتقدير الكثافة سيكون

$$\hat{D} = \frac{y\hat{f}(0)}{2L}$$

والمشكلة التي تواجهنا تمكن بتقدير $f(0)$ الكثافة عند المسافة صفر من الخط القاطع.

يمكننا أن نتصور شريطاً مماثلاً من الأرض بعرض w تكون فيه قابلية الاكتشاف كاملةً، ويكون عدد الحيوانات المكتشفة في هذا الشريط متساوية بالمعدل لعدد الحيوانات المكتشفة من الخط القاطع، وتكون قابلية الاكتشاف متناقصة مع المسافة، العلاقة بين الخط القاطع والشريط المماثل من الأرض تكون

$$f(0) = \frac{1}{w}$$

وw يدعى نصف العرض الفاعل للخط القاطع. وبدلالة w يمكن تقدير الكثافة بالاعتماد على w الذي هو تقدير إلى w لتكون

$$\hat{D} = \frac{y}{2Lw}$$

يمكننا إما تقديرw أو $f(0)$ ومن ثم نستطيع تقدير الكثافة.

عندما يمكننا افتراض توزيع احتمالي معين يعتمد على معالم غير معلومة لدالة الاكتشاف $g(x)$، نستطيع أن نستخدم بعض الطرق الإحصائية مثل طريقة التقدير الاحتمالي العظمى (mle) لتقدير المعالم غير المعلومة من ثم نقدر $f(0)$ أو w. لقد تم دراسة بعض التوزيعات الاحتمالية التي تعتمد على معالم غير معلومة من قبل مجموعة من الباحثين منهم: Ramsey(1979)

و Buckland(1985) و Burnham et al. (1980) و Qiunn and Gallucci(1980) و Pollock(1978). سوف نستخدم توزيعان من أبسط التوزيعات كأمثلة في هذا الفصل.

لابد من الإشارة إلى أن من الفوائد لافتراض توزيع بسيط لدالة الاكتشاف سيؤدي إلى تقدير بسيط لكثافة المجتمع إذا كان الافتراض صحيحاً، ولكن من عيوبه أنه قد لا يكون لهذا التوزيع البسيط المرونة الكافية لتمثيل دالة الاكتشاف الحقيقية. سنتناول هنا توزيعين لدالة الاكتشاف هما التوزيع الأسي والتوزيع الطبيعي النصفي؛ وذلك لكونهما يؤديان إلى تقديرات بسيطة للكثافة.

في حالة التوزيع الأسي تكون دالة الاكتشاف $g(x)=\exp(-x/w)$. كلما كانت قيمة المعلمة w كبيرة أدى إلى أن تكون قابلية الاكتشاف من الخط كبيرة. باستخدام طريقة التقدير الاحتمالي العظمى (mle) نحصل على تقدير إلى w هو $w=\bar{x}$ أي معدل مسافات الاكتشاف. لمزيد من المعلومات يراجع Ramsey(1979).

مثال 2: لنفترض أنَّ دالة الاكتشاف للبيانات في المثال 1 تتبع التوزيع الأسي أي $g(x)=\exp(-x/w)$. قدر كثافة الطيور في منطقة الدراسة.

الحل: باستخدام البيانات في المثال السابق نستطيع أن نحصل على $w=\bar{x}=16.39$ وعليه فسيكون تقدير الكثافة

$$\hat{D}=\frac{y}{2Lw}=\frac{18}{2(100)(16.39)}=0.0055$$

بالرغم من كون استخدام التوزيع الأسي يؤدي إلى تقدير بسيط جداً لكثافة المجتمع، ولكنه غير واقعي لكثير من المجتمعات ولا يؤدي إلى تقدير دقيق للكثافة. أكثر من باحث مثل Buckland (1985) و Burnham et al.

(1980) و(1978) Eberhardt بحثوا هذا الموضوع وقالوا: أن دالة الاكتشاف يجب أن يكون لها مشتقة تساوي صفراً قريباً من الخط القاطع. أبسط التوزيعات التي ينطبق عليه هذا الشرط هو التوزيع الطبيعي النصفي. وبدالة احتمال هي

$$g(x) = \exp\left(\frac{-\pi x^2}{4w^2}\right)$$

وبتقدير إلى w

$$w = \sqrt{\frac{\pi}{2y}\sum_{i=1}^{y} x_i^2}$$

مثال 3: لنفترض أن دالة الاكتشاف للبيانات في المثال 1 تتبع التوزيع الطبيعي النصفي. قدر كثافة الطيور في منطقة الدراسة.

الحل: باستخدام البيانات في المثال 1 يمكننا حساب

$$\frac{1}{y}\sum_{i=1}^{y} x_i^2 = \frac{1}{18}(0^2 + 0^2 + \cdots + 44^2) = 417.5$$

لذا فإن تقدير w سيكون

$$w = \sqrt{\frac{\pi}{2y}\sum_{i=1}^{y} x_i^2} = \sqrt{\left(\frac{3.1417}{2}\right)(417.5)} = 25.61$$

وعليه فإن تقدير الكثافة سيكون

$$\hat{D} = \frac{y}{2Lw} = \frac{18}{2(100)(25.6)} = 0.0035$$

3.2.15 الطرق غير المعلماتية (Nonparametric Methods)

لتفادي افتراض أن دالة الاكتشاف تتبع توزيعاً احتمالياً معيناً يمكننا أن نستخدم الطرق غير المعلماتية للتقدير، باستخدام البيانات لمتغير عشوائي من

دالة كثافة الاحتمال f، الطرق غير المعلماتية تستخدم طرق تمهيد smoothing معينة لتقدير قيمة $f(x)$ دالة كثافة الاحتمال لأي قيمة للمتغير x. وللخط القاطع فإن دالة كثافة الاحتمال التي تهمنا هي التي الدالة تقيس وتكتشف المسافات من الخط. سنتناول طريقتين من الطرق غير المعلماتية في هذا الفصل، ولمزيد من المعلومات يراجع Buckland(1985) و .Burnham et al (1980) و(1985) Johnson and Routledge. قابلية الاكتشاف على الخط تكون كاملة، لذا فإن تقدير الكثافة يكون $\hat{D} = y\hat{f}(0)/2L$.

تقدير $f(0)$ باستخدام طريقة الكرنل (Kernel Method)

هناك كم كبير من البحوث الإحصائية التي تتناول تقدير دالة كثافة الاحتمال معظمها استخدم طريقة الكرنل للتقدير. يُعَدُّ Seber (1986) أول من اقترح استخدام طريقة الكرنل لتقدير الكثافة باستخدام الخط القاطع، وقد استخدمت الطريقة من قبل Quang (1993) لمسائل قريبة.

تستخدم الطريقة دالة الكرنل $K(x)$ التي يكون تكاملها 1

$$\int_{-\infty}^{\infty} K(x)\,dx = 1$$

إن تقدير دالة كثافة الاحتمال f عند القيمة x باستخدام طريقة الكرنل هو

$$\hat{f}(x) = \frac{2}{yh}\sum_{j=1}^{y} K\left(\frac{x - x_j}{h}\right)$$

حيث إن h تمثل عرض النافذة وx_j قيمة المشاهدة j (التي تمثل المسافة العمودية من الخط القاطع إلى الحيوان j) وy تمثل عدد الحيوانات التي تم مشاهدتها، أما الرقم 2 فيستخدم عندما تحسب الكثافة بغض النظر عن أي جانب من الخط الذي تمت مشاهدة الحيوان عنده.

يمكننا تقدير $f(0)$ مع كون دالة الكرنل متماثلة بما يأتي

$$\hat{f}(0) = \frac{2}{yh} \sum_{j=1}^{y} K\left(\frac{x_j}{h}\right)$$

فإذا كانت دالة الكرنل تتبع التوزيع الطبيعي فإن

$$k\left(\frac{x_j}{h}\right) = \frac{1}{\sqrt{2p}} e^{-(1/2)(x_j/h)^2}$$

لقد أعطى Silveran (1986) قاعدة بسيط الاختيار عرض النافذة h وهي

$$h = 0.9 a y^{-1/5}$$

حيث إن $a=\min(s,Q/1.34)$ ، أما s فتمثل الانحراف المعياري للمشاهدات التي هي عبارة عن المسافة العمودية للحيوانات المكتشفة من الخط، وQ فتمثل المدى الربيعي، ولكن إذا كنا نتعامل مع مسافة موجبة فقط فيجب أن نستخدم الوسيط بدلاً من المدى الربيعي.

تقدير كثافة الحيوانات في المجتمع ستكون

$$\hat{D} = \frac{y\hat{f}(0)}{2L}$$

مثال 4: إذا كانت دالة الكرنل تتبع التوزيع الطبيعي، استخدمت البيانات في المثال 1 لتقدير كثافة الطيور في منطقة الدراسة.

الحل: الوسيط للقيمة المطلقة للبيانات في المثال 1 هو 15 و$15/1.34=11.19$. بما أن 11.19 أقل من قيمة الانحراف المعياري $s=12.56$ وباستخدام الطريقة التي اقترحها Silveran (1986) يمكننا إيجاد عرض النافذة

$$h = 0.9(11.19)(18)^{-1/5} = 5.56$$

لذا سيكون تقدير $f(0)$ باستخدام طريقة الكرنل مع دالة تتبع التوزيع الطبيعي هو

$$\hat{f}(0) = \frac{2}{yh}\sum_{j=1}^{y} K\left(\frac{x_j}{h}\right) = \frac{2}{yh}\sum_{j=1}^{y}\frac{1}{\sqrt{2\pi}}\, e^{-(1/2)(x_j/h)^2}$$

الآن باستخدام البيانات في المثال 1 نستطيع تقدير $f(0)$

$$\hat{f}(0)=\frac{2}{yh}\sum_{j=1}^{y}\frac{1}{\sqrt{2p}}\, e^{-(1/2)(x_j/h)^2} = \frac{2}{18(5.65)\sqrt{2p}}\left[e^{(-1/2)(0/5.65)^2} + L + e^{(-1/2)(44/5.65)^2}\right]=0.0376$$

يمكننا استخدام $\hat{f}(0)$ لتقدير كثافة الطيور في منطقة الدراسة

$$\hat{D} = \frac{y\hat{f}(0)}{2L} = \frac{18(0.0376)}{2(100)} = 0.0034$$

تقدير $f(0)$ باستخدام طريقة سلاسل فورير (Fourier Series Method)

إن تقدير $f(0)$ باستخدام طريقة سلاسل فورير هو

$$\hat{f}(0) = \frac{1}{w^*} + \sum_{k=1}^{m}\hat{A}_k$$

حيث إن w^* يمثل المسافة العظمى التي يمكن أن يشاهد فيها الحيوان و $\hat{A}_k$ يمكن تعريفه

$$\hat{A}_k = \frac{2}{yw^*}\left[\sum_{i=1}^{y}\cos(k\pi x_i/w^*)\right]$$

أما m فتمثل عدد الحدود التي سنستخدمها لإيجاد القيم التقريبية إلى $\hat{f}(0)$ ، ويمكننا استخدام القاعدة التي اقترحها Burnham et al. (1980) التي تقول ابدأ بـ $m=1$ ومن ثم اختر أول عدد صحيح m بحيث إن

$$\frac{1}{w^*}\sqrt{2/(y+1)} \ge \left|\hat{A}_{m+1}\right|$$

أما فيما يخص تحديد w^* المسافة العظمى التي يمكن أن يشاهد فيها الحيوان فقد اقترح كلٌّ من (1980) .Burnham et al و(1979) .Crain et al استخدام مسافة أقل من المسافة العظمى التي تمت مشاهدة أحد الحيوانات فيها ، وذلك من خلال إهمال 1 إلى 3% من المشاهدات الكبيرة على أنها شاذة. لمزيد من المعلومات يراجع (1981) .Burnham et al و(1990) Quang.

مثال 5: استخدم طريقة سلاسل فورير لتقدير $f(0)$ واستخدمها لتقدير كثافة الطيور في منطقة الدراسة.

الحل: إذا استخدمنا البيانات في المثال 1 فإن القيمة العظمى هي 44 متراً ، وإذا أهملنا هذه القيمة على أنها قيمة شاذة فإننا سوف نستخدم $w^*=34$ والقاعدة أعلاه تكون متحققة عندما تكون $m=1$، لذلك نحتاج إلى حد واحد فقط، لذلك نحتاج حساب $\hat{A}_1$ فقط. وهو الذي يحتوي على17 حداً ، تمثل عدد المشاهدات بعد إهمال القيمة العظمى.

$$\hat{A}_1=\frac{2}{17(34)}[\cos(\frac{1(3.14 17)(0)}{34})+L+\cos(\frac{1(3.1417)(34)}{34})]=0.0091$$

لذا فإن تقدير $f(0)$ سيكون

$$\hat{f}(0)=\frac{1}{w^*}+\sum_{k=1}^{m}\hat{A}_k=\frac{1}{34}+0.0091=0.0385$$

أخيراً فإن تقدير الكثافة يكون

$$\hat{D}=\frac{y\hat{f}(0)}{2L}=\frac{18(0.0385)}{2(100)}=0.0033$$

3.15 تصاميم لاختيار الخطوط

إن تصميم المعاينة للمسح بالخط القاطع عبارة عن الطريقة التي يتم فيها اختيار مواقع الخطوط، تعتمد بعض الخواص المهمة مثل عدم الانحياز لتقدير

الكثافة أو تقدير التباين على تصميم الخطوط وليس على الفروض حول مجتمع الدراسة.

يمكن أن نشاهد مجموعة كبيرة من الحيوانات باستخدام خط واحد، ولكن لا يزال حجم العينة 1. وللحصول على تقدير دقيق للكثافة لابد من حجم عينة مساوٍ إلى n من الخطوط، بصورة خاصة إذا كان توزيع الحيوانات في المجتمع غير متماثل في مناطقه المختلفة.

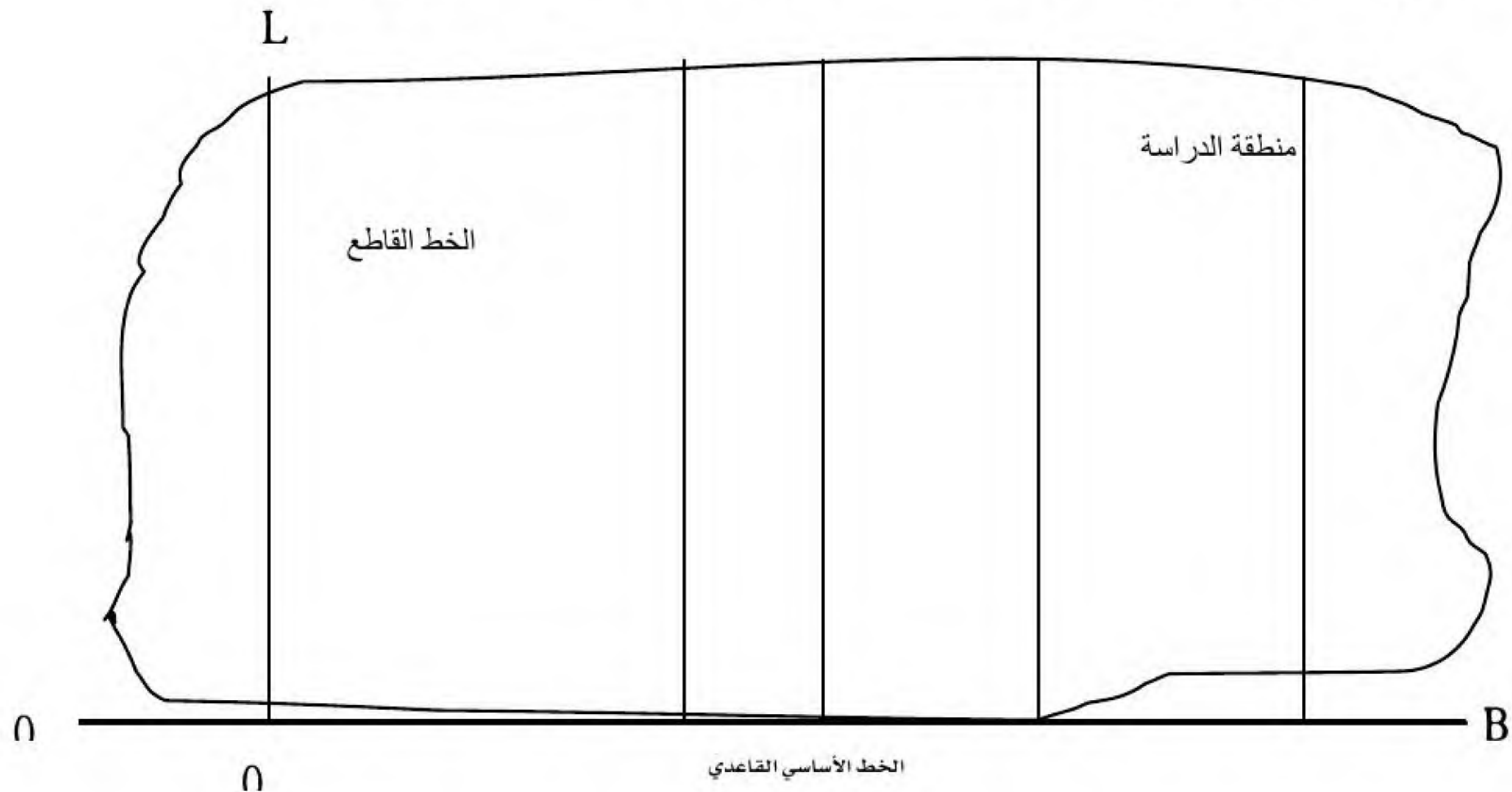
إن تقدير التباين بالاعتماد على عينة من عدة خطوط يفضل على التقدير الذي يعتمد على خط واحد. ولقد أكد وأوصى عدد كثير من الباحثين على ضرورة أن يعتمد تقدير التباين على أكثر من خط مثل: Burnham and Anderson (1976) وEberhardt(1978) وOverton (1969) وSeber(1982). تعتمد الطرق التي تقدر التباين باستخدام خط واحد على وضع فروض على توزيع الحيوانات بالمجتمع منها: أن الحيوانات تتوزع بشكل متماثل وبصورة مستقلة في منطقة الدراسة، ولابد من الإشارة هنا إلى أن مثل هذه الفروض غير عملية لأنَّ الحيوانات غالباً ما تعيش في مجتمعات مع بعضها بعضًا في أجزاء معينة من مجتمعات وجودها.

4.1.5 اختيار الخطوط بطريقة عشوائية

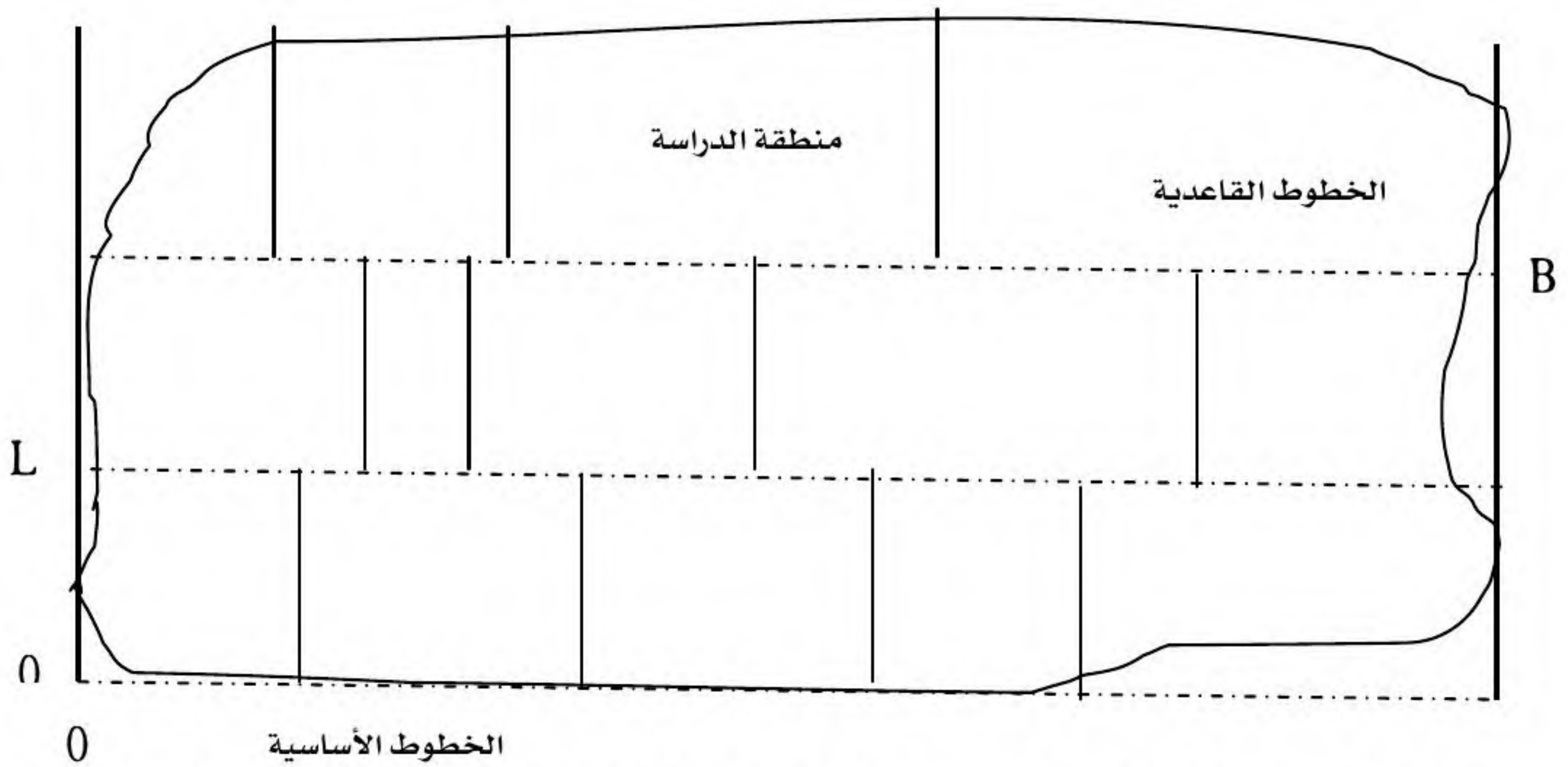
يمكن اختيار عينة عشوائية بحجم n من الخطوط في منطقة الدراسة وذلك برسم الخط الأساسي بطول B على خريطة منطقة الدراسة خلال منطقة الدراسة أو تحتها كما هو موضح في الشكل 3. ليس بالضرورة أن تكون منطقة الدراسة منتظمة الشكل. بعد ذلك نقوم بسحب n من الأعداد العشوائية من التوزيع المنتظم التي تقع في الفترة بين 0 وB ولتكن $v_1, v_2, ..., v_n$ ومن ثم نقوم برسم n من خطوط التقاطع التي تكون مواضعها $v_1, v_2, ..., v_n$ وتكون عمودية على الخط الأساسي. الشكل 3

يحتوي على خمسة خطوط عمودية على الخط الأساسي بطول L متر جرى تحديد مواضعها بسحب $v_1, v_2, ..., v_5$ من الأعداد العشوائية في الفترة بين 0 وB.

إذا كانت منطقة الدراسة عرضها أكبر من طول الخط القاطع، نقوم برسم خطوط أساسية متوازية تبعد عن بعضها بعضًا مسافة L متر، ومن ثم نقوم بسحب نقاط البداية $v_1, v_2, ..., v_n$ للخطوط القاطعة في الفترة بين 0 وB كما هو موضح بالشكل4، تمثل الخطوط المتقطعة الخطوط الأساسية، وهنالك 11 خطاً قاطعاً جرى تحديد مواضعها في الفترة 0 وB بسحب نقاط البداية $v_1, v_2, ..., v_{11}$ من الأعداد العشوائية في الفترة بين 0 وB.



الشكل3: عينة عشوائية من 5 خطوط بطول L متر لكل خط في منطقة الدراسة. إذا كانت منطقة الدراسة عرضها أكبر من طول الخط القاطع، فعلينا أن نقوم برسم خطوط أساسية متوازية تبعد عن بعضها الآخر بمسافة L متر. ومن ثم نقوم بسحب نقاط البداية $v_1, v_2, ..., v_n$ للخطوط القاطعة في الفترة بين 0 وB كما هو موضح بالشكل 4. تمثل الخطوط المتقطعة الخطوط الأساسية، وهنالك 11 خطاً قاطعاً جرى تحديد مواضعها في الفترة 0 وB بسحب نقاط البداية $v_1, v_2, ..., v_{11}$ من الأعداد العشوائية في الفترة بين 0 وB.



الشكل4: عينة عشوائية من 11 خطوط بطول L متر لكل خط في منطقة الدراسة

قد تكون منطقة الدراسة غير منتظمة لهذا لا يمكن أن نحصل على خطوط متساوية الأطوال، أي أن طول الخط القاطع يكون متغيراً عشوائياً L_i. لمزيد من المعلومات يراجع Thompson (2002).

5.15 اختيار الخطوط باحتمالات متناسبة مع أطوالها

يمكننا اختيار الخطوط باحتمالات متناسبة مع أطوالها وذلك عن طريق اختيار n من النقاط بصورة مستقلة من التوزيع المنتظم من منطقة الدراسة. ويمكن تحقيق ذلك من خلال وضع منطقة الدراسة داخل مستطيل، من ثم نقوم باختيار إحداثيات النقاط بطريقة عشوائية حتى نحصل على n من النقاط داخل منطقة الدراسة. لكل نقطة جرى سحبها داخل منطقة الدراسة نقوم برسم خط يمر بهذه النقطة، ويكون عمودياً على الخط الأساسي ويقطع منطقة الدراسة. نلاحظ أن الخط الذي يقع في جزء واسع من منطقة الدراسة يكون طويلاً، لذا فإن احتمال اختياره يكون أكبر؛ لأنَّ هنالك عدد أكبر

من النقاط في منطقة الدراسة تؤدي إلى اختياره؛ لذا فإن دالة كثافة الاحتمال للخط القاطع الذي بدأ من الموضع v على الخط الأساسي ستكون $L(v)/A$.

إذا كان لدينا خط واحد جرى سحبه بالطريق أعلاه، يمكننا أن نقدر الكثافة بما يأتي

$$\hat{D} = \frac{y}{2wL(v)}$$

حيث إن w تأثير نصف العرض للخط القاطع، وv نقطة التقاطع للخط القاطع مع الخط الأساسي، و$L(v)$ عرض منطقة الدراسة أو طول الخط العمودي من نقطة التقاطع الواقعة على الخط الأساسي إلى نهاية منطقة الدراسة. بعد هذا يمكننا إثبات أن $\hat{D}$ تقدير غير متحيز إلى D، لمزيد من المعلومات يراجع(2002) Thompson.

لنفترض أن $\hat{D}_i$ تمثل تقدير الكثافة من الخط i في العينة بالحجم n، لذا فإن تقدير الكثافة سيكون

$$\hat{D}_p = \frac{1}{n}\sum_{i=1}^{n}\hat{D}_i$$

هو تقدير غير متحيز إلىD. وبما أن الخطوط سحبت بطريقة عشوائية بالاعتماد على نقاط البداية، فإن $\hat{D}_i$ تُعَدُّ متغيرات عشوائية ومستقلة ولديها التوزيع نفسه. ويمكننا أن نحصل على تقدير غير متحيز لتقدير تباين $\hat{D}_p$ باستخدام $\hat{D}_i$ وهو

$$s^2(\hat{D}_p) = \frac{1}{n(n-1)}\sum_{i=1}^{n}(y_i - \hat{D}_p)^2$$

References

1. Barry, S. C. and Welsh, A. H. (2001). Distance Sampling Methodology, *Journal of the Royal Statistical Society B*. 63, 31–51.
2. Buckland, S. T. (1982). A Note on the Fourier Series Model for Analysing Line Transect Data, *Biometrics*. 38, 469–477.
3. Buckland, S. T. (1985). Perpendicular Distance Models for Line Transect Sampling, *Biometrics*. 41, 177–195.
4. Buckland, S. T. (1987). One the Variable Circular Plot Method of Estimating Animal Density, *Biometrics*. 43, 363–384
5. Buckland, S. T., Anderson, D. R., Burnham, K.P. and Laake, J. L. (1992). Distance Sampling Estimating of Abundance of Biological Population, Chapman & Hall, London.
6. Burnham, K.P. (1979). A Parametric Generalization of the Hayne Estimator for Line Transect Sampling, *Biometrics*. 35, 587–595.
7. Burnham, K.P. and Anderson, D. R (1976). Mathematical Models foe Nonparametric Inferences from Line Transect Data, *Biometrics*. 32, 325–336.
8. Burnham, K.P., Anderson, D. R. and Laake, J. L. (1980). Estimation of Density from Line Transect Sampling of Biological Populations, *Wildlife Monograph*. 72, Supplement to the Journal of Wildlife Management.
9. Burnham, K.P., Anderson, D. R. and Laake, J. L. (1981). Line Transect Estimation of Bird Population Density using Fourier Series, *Studies in Avian Biology*. 6, 466–482.
10. Crain B. R., Burnham, K.P., Anderson, D. R. and Laake, J. L. (1979). Nonparametric Estimation of Population Density for Line Transect Sampling using Fourier Series, *Biometrical Journal*. 21, 731–748.
11. Drummer, T. D. and McDonald, L. L. (1987). Size Biased in Line Transect Sampling, *Biometrics*. 43, 13–21.
12. Eberhardt, L. L. (1978). Transect Methods for Population Studies, *Journal of Wildlife Management*. 42, 1–31.
13. Gates, C. E. (1979). Line Transect and Related Issues. In R. M. Cormack, G. P. Patil and D. S. Robson (eds), Sampling Biological Population. Fairland, MD: International Co-operative Publishing House, 71–154.
14. Johnson, E. G. and Routledge, R. D. (1985). The Line Transect Method: A Nonparametric Estimator based on Shape Restrictions, *Biometrics*. 41, 669–679.
15. Otto, M. C. and Pollock, K. H. (1990). Size Biased in Line Transect Sampling: A Field Test. *Biometrics*. 46, 239–245.
16. Overton, W. S. (1969). Estimating the number of Animals in Wildlife Populations: In R. H. Giles (eds.) *Wildlife Management Techniques*, 3rd ed. Washington, DC: Wildlife Society, 403–455.
17. Pollock, K. H. (1978). A Family of Density Estimators for Line Transect Sampling, *Biometrics*. 34, 475–478.
18. Quang, P. X. (1990). Confidence Intervals for Densities in Line Transect Sampling, *Biometrics*. 46, 497–472.

19. Quang, P. X. (1991). A Nonparametric Approach to Size-biased Line Transect Model for Aerial Surveys, *Biometrics*. 47, 269-279.
20. Quang, P. X. and Lanctot, R. B. (1991). A Line Transect Sampling, *Biometrics*. 47, 1089-1102.
21. Quinn, T. J., II and Gallucci, V. F. (1980). Parametric Models for Line Transect Estimators of Abundance, *Ecology*. 61, 293-302.
22. Ramsey, F. L. (1979). Parametric Models for Line Transect Surveys, *Biometrika*. 66, 505-512.
23. Ramsey, F. L., Gates, C. E., Patil, G. P. and Taillie, C. (1988). On Transect Sampling to Assess Wildlife Populations and Marine Resources. In P. R. Krishnaiah and C. R. Rao (eds.), Handbook of Statistics, Vol. 6 Sampling. Amsterdam: Elsevier Sciences Publishers, 515-532.
24. Schweder, T. (1977). Point Process Models for Line Transect Experiments. In J. R. Barra, F. Brodeau, G. Romier, and B. Van Cutsem (eds.), Recent Development in Statistics. Amsterdam: North-Holland, 221-242.
25. Seber, G. A. F. (1982). The Estimation of the Animal Abundance, 2nd ed. London, Griffin.
26. Seber, G. A. F. (1986). A Review of Estimating Animal Abundance. *Biometrics*, 42, 267-292.
27. Seber, G. A. F. (1992). A Review of Estimating Animal Abundance. *II International Statistical Review*, 60, 129-166.
28. Silverman, B. W. (1986). Density Estimation for Statistics and Data Analysis, Chapman & Hall, London.

الفصل السادس عشر

المعاينة بخط التقاطع

Line Intercept sampling

1.16 مقدمة

إن المعاينة بطريقة خط التقاطع عبارة عن عينة من الخطوط يجري اختيارها في منطقة الدراسة، وعندما يقطع خط أو أكثر أحد عناصر المجتمع نقوم بقياس متغير أو أكثر في هذا العنصر. على سبيل المثال ممكن أن تكون الشجيرات المنتشرة في منطقة الدراسة هي العناصر التي نريد دراستها، فعندما يقطع خط أو أكثر إحدى هذه الشجيرات نقوم بقياس –على سبيل المثال– عدد الأغصان بهذه الشجيرة أو ارتفاعها أو عرضها أو عدد الثمرات الموجدة فيها ... إلخ.

إن الشجيرات الكبيرة في منطقة الدراسة يكون احتمال أن يقطعها خط أو أكثر أكبر من الشجيرات الصغيرة، لذا فإن الحصول على تقديرات غير متحيزة لمعالم المجتمع تعتمد على تحديد هذه الاحتمالات التي تعتمد على حجم الوحدات.

فيما يأتي بعض الرموز والمصطلحات التي سوف نستخدمها في هذا الفصل

A : مساحة منطقة الدراسة.

a_i : مساحة العنصر i.

y_i : صفة ما نريد قياسها في العنصر i (إنتاج العنصر أو ارتفاعه أو عدد الأغصان).

N : عدد العناصر في منطقة الدراسة (16 = N في الشكل 1).

$\tau_y = \sum_{i=1}^{N} y_i$: المجموع الكلي للمجتمع للمتغير y.

μ_y : الوسط الحسابي للمجتمع للمتغير y.

$C = \sum_{i=1}^{N} a_i / A$: تغطية N من العناصر لمنطقة الدراسة.

B: طول الخط الأساس.

L: طول الخط القاطع.

v_i : طول التقاطع بين خط التقاطع والعنصر i كما هو موضح بالشكل 1.

w_i : تمثل ضل (عرض) العنصر i على خط الأساس كما هو موضح بالشكل 1

n: عدد العناصر التي قُطعت بخط التقاطع.

m: حجم العينة وهو يمثل عدد الخطوط التي ثم أختيرت في منطقة الدراسة.

$D=N/A$ كثافة العناصر في منطقة الدراسة.

216 عينة عشوائية من خطوط ثابتة الاتجاه

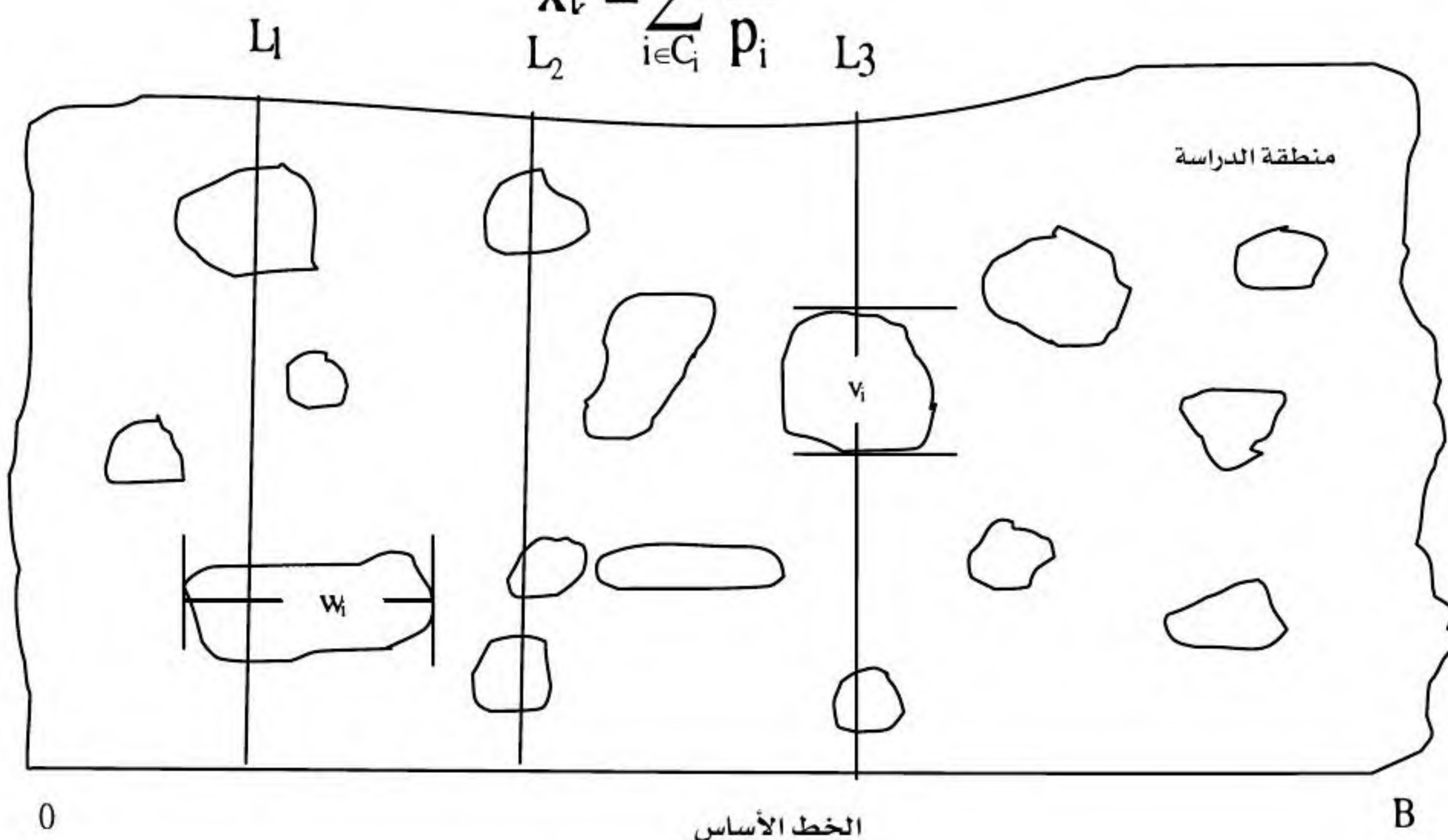
إن من أبسط التصاميم هو اختيار m من الخطوط بطريقة عشوائية وذلك باختيار m من المواضع بطريقة عشوائية على طول الخط الأساس لمنطقة الدراسة والذي يمثل عرض منطقة الدراسة. ومن ثم نقوم بمد الخطوط القاطعة التي تكون عمودية على الخط الأساس. الشكل 1 يوضح هذا التصميم. في الشكل 1 لدينا ثلاثة خطوط متساوية الطول.

لنفترض أن لدينا N من العناصر في منطقة الدراسة، مع كل عنصر لدينا متغير واحد على الأقل نرغب في دراسته وليكن y، الهدف هو تقدير مجموع المجتمع الكلي $\tau = \sum_{i=1}^{N} y_i$ للمتغير y أو الكثافة لكل وحدة مساحة $D_y = \tau / A$. إن احتمال أن يقطع الخط k العنصر i يكون متناسباً مع عرض العنصر w_i كما هو موضح في الشكل 1؛ لذا فإن احتمال أن يقطع الخط k العنصر i هو

$$p_i = \frac{w_i}{B}$$

لنفترض أن C_i تمثل مجموعة العناصر التي قطعت من قبل الخط k في عينة حجمها m من الخطوط. لنعرف متغيراً جديداً هو

$$x_k = \sum_{i \in C_i} \frac{y_i}{p_i}$$



الشكل 1: عينة عشوائية من 3 خطوط ثابتة الاتجاه بطول L لكل خط في منطقة الدراسة

نلاحظ أن المتغير x_k عبارة عن تقدير غير متحيز إلى المجموع الكلي للمجتمع τ. وبما أن العينة العشوائية التي تتكون من m خط سوف تعطينا $x_1, x_2, ..., x_m$ –هذه التقديرات تتوزع بشكل مستقل ومشابه لبعضها الآخر– فإن وسطها الحسابي يمثل تقدير غير متحيز إلى τ

$$\hat{\tau} = \frac{1}{m}\sum_{k=1}^{m} x_k$$

وبما أن العناصر لم تُسحب بصورة مستقلة عن بعضها الآخر –حيث إن العنصر يمكن أن يُقطع من قبل أكثر من خط– فلابد من تكرار الخطوط أو بعبارة أخرى سَحْب عينة عشوائية بحجم m من الخطوط لنستطيع أن نستخدمها للحصول على تقدير غير متحيز إلى تباين $\hat{\tau}$ باستخدام

$$s^2(\hat{\tau}) = \frac{s_x^2}{m}$$

حيث إن

$$s_x^2 = \frac{1}{m-1}\sum_{k=1}^{m}(x_k - \hat{\tau})^2$$

3.16 عينة من خطوط عشوائية الموضع والاتجاه

لنفترض أن كل خط جرى سحبه بطريقة عشوائية كاملة من منطقة الدراسة. يمكننا أن ننفذ هذه الطريقة وذلك بأن نسحب مجموعة من النقاط بعدد الخطوط التي ننوي سحبها بطريقة عشوائية من منطقة الدراسة، كل نقطة تمثل نقطة المنتصف لكل خط طوله L. ومن هنا نقوم بتحديد اتجاه

الخط باختيار قيمة الزاوية بطريقة مستقلة، ولتحديد قيمة الزاوية نقوم بسحب رقم بطريقة عشوائية من التوزيع المتماثل في الفترة $[0, \pi)$. وإذا حدث أن خرج قسم من الخط خارج منطقة الدراسة فعلينا أنْ نقوم بمد هذا القسم في منطقة أخرى داخل منطقة الدراسة. لمزيد من المعلومات راجع Kaiser (1983).

بمعرفة اتجاه الخط θ فإن احتمال أن يكون العنصر i قطع من قبل الخط هو

$$P_i(\theta) = L\, w_i(\theta)/A$$

حيث أن $w_i(\theta)$ عبارة عن عرض العنصر باتجاه عمودي على θ أي المسافة القصوى بين خطين باتجاه θ التي قطعت العنصر كما هو مبين بالشكل رقم2.

أما الاحتمال غير المشروط لاختيار العنصر i فهو

$$p_i = LC_i/A$$

حيث إن

$$C_i = E[w_i(\theta)]$$

أي توقع $w_i(\theta)$ لتوزيع θ.

يمكننا أن نجد تقدير غير متحيز للمجموع الكلي للمجتمع τ باستخدام الاحتمال المشروط أو الاحتمال غير المشروط، يمكننا أن نعرف متغير جديد للخط k

$$x_k(\theta) = \sum_{i \in C_i} \frac{y_i}{P_i(\theta)}$$

أو

$$x_k = \sum_{i \in C_i} \frac{y_i}{p_i}$$

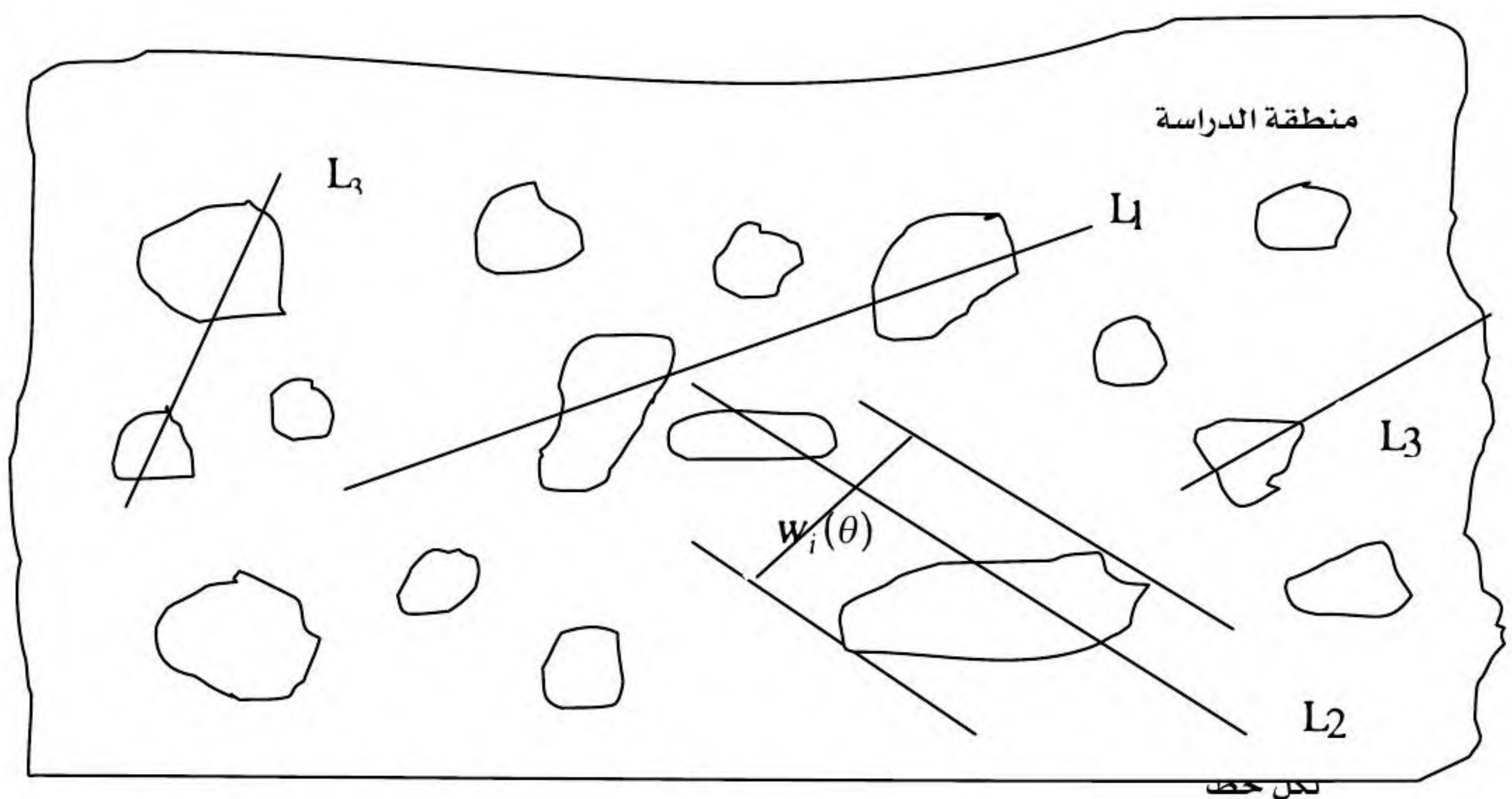
حيث أن كلاً من $x_k(\theta)$ و x_k تقديرين غير متحيزين إلى τ.

وإذا قمنا بسحب عينة بحجم m من الخطوط فسنحصل على تقديرين غير متحيزين إلى τ هما

$$\hat{\tau}_{p(\theta)} = \frac{1}{m} \sum_{k=1}^{m} x_k(\theta)$$

و

$$\hat{\tau}_p = \frac{1}{m} \sum_{k=1}^{m} x_k$$



الشكل2: عينة عشوائية من 3 خطوط متغيرة الاتجاه والموضع بطول L

ويمكننا أن نحصل على تقدير غير متحيز لتقدير التباين لكلا التقدير باستخدام تباين العينة لكلٍ من $x_k(\theta)$ و x_k وهما على التوالي

$$s^2(\hat{t}_{p(\theta)}) = s^2_{x(\theta)}/m$$

حيث إن

$$s^2_{x(\theta)} = \frac{1}{m-1}\sum_{k=1}^{m}(x_k(\theta) - \hat{t}_{p(\theta)})^2.$$

و

$$s^2(\hat{t}_p) = s^2_x/m$$

حيث إن

$$s^2_x = \frac{1}{m-1}\sum_{k=1}^{m}(x_k - \hat{t}_p)^2$$

ولمقارنة أي التقديرين $\hat{t}_{p(\theta)}$ و $\hat{t}_p$ أفضل يراجع Kimura and Lemberg (1981).

ولابد من الإشارة هنا إلى أن استخدام الخطوط العشوائية (الموضع والاتجاه) يوفر لنا تقديرات غير متحيزة لتقدير المجموع الكلي (الكثافة والتغطية) ولكن علمية تنفيذها في حقل الدراسة قد تكون صعبة إن لم تكن مستحيلة إذا كان حقل الدراسة كبيرًا جداً، لذا نوصي باستخدام طريقة الخطوط الثابتة الاتجاه التي تكون عمودية على خط الأساس، أو استخدام العينة المنتظمة من الخطوط العمودية على خط الأساس. لمزيد من المعلومات يراجع Lucsa and Seber (1977) وMcDonald(1980).

4.16 تقدير التغطية والكثافة باستخدام خطوط التقاطع

إن من أهم استخدامات معاينة خطوط التقاطع هو تقدير التغطية والكثافة لعنصر أو أكثر في منطقة الدراسة. إن وحدة المعاينة هي خط التقاطع الذي يكون عمودياً على الخط الأساس ويقطع منطقة الدراسة، أي يكون طوله بعرض منطقة الدراسة. إذا كانت منطقة الدراسة مستطيلة فسيكون طول الخطوط متساوياً وبخلاف ذلك ستكون أطوال الخطوط غير متساوية. وبما أن العناصر في منطقة الدراسة لن تكون متساوية الحجم، لذا فإن احتمالات أن قطع هذه العناصر من قبل الخط القاطع غير متساوية، وستكون هذه الاحتمالات متناسبة مع حجم العناصر في منطقة الدراسة.

1.4.16 تقدير التغطية

لقد استُخدمت خطوط التقاطع لتقدير التغطية لعنصر أو عناصر معينة لمنطقة الدراسة أول مرة من قبل Canfield (1941). من البدهي أن تقدر نسبة التغطية لعنصر معين في منطقة الدراسة بما يغطيه هذا العنصر لخط. إذا كان لدينا خط واحد بطول L فيمكننا أن نقدر التغطية باستخدام

$$\hat{C} = \sum_{i=1}^{n} v_i / L$$

يمكننا أن نثبت أن هذا التقدير هو تقدير غير متحيز إلى التغطية C، يراجع Lucas and Seber (1977). لتقدير التباين لابد أن نضع افتراضات معينة على توزيع العناصر في المجتمع، وفي معظم الأحيان هذه الافتراضات غير واقعية، أو نقوم بسحب عينة عشوائية من الخطوط.

لنفترض أننا قمنا بسحب عينة بحجم m من الخطوط وبطول L نفسه. وقمنا بتقدير التغطية لكل خط وهي

$$\hat{c}_k = \sum_{i=1}^{n_k} v_i / L, \quad k = 1, 2, \ldots, m$$

حيث إن n_k يمثل عدد العناصر التي تم قطعها من قبل الخط k. وبما أن العينة العشوائية التي تتكون من m خط سوف تعطينا تقديرات $\hat{c}_1, \hat{c}_2, \ldots, \hat{c}_m$ وهذه التقديرات تتوزع بشكل مستقل ومشابه لبعضها الآخر؛ لذا فإن وسطها الحسابي يمثل تقديرًا غير متحيز إلى C وهو

$$\hat{c} = \frac{1}{m} \sum_{k=1}^{m} \hat{c}_k$$

يمكن أن نحصل على تقدير غير متحيز إلى تقدير تباين $\hat{c}$ باستخدام

$$s^2(\hat{c}) = \frac{s_c^2}{m}$$

حيث إن

$$s_c^2 = \frac{1}{m-1} \sum_{k=1}^{m} (\hat{c}_k - \hat{c})^2$$

أما إذا كانت أطوال الخطوط غير متساوية فيجب أن نستخدم الوسط الحسابي المرج لتقدير قيمة التغطية وهو

$$\hat{c} = \sum_{k=1}^{m} L_k \hat{c}_k / \sum_{k=1}^{m} L_k$$

2.4.16 تقدير الكثافة

لنفترض أن منطقة الدراسة عبارة عن مستطيل بمساحة $L \times B$ كما هو موضح بالشكل 1 أعلاه. وفي هذا المستطيل نقوم بمد خط التقاطع وذلك بعد

تحديد موضعه على خط الأساس باختيار نقطة عشوائية داخل الفترة $(0, B)$ ومن ثم مد الخط ليقطع منطقة الدراسة ويكون عمودياً على الخط الأساس. إن احتمال أن يقطع الخط العنصر i يكون متناسباً مع عرض العنصر w_i وهو

$$p_i = \frac{w_i}{B},\ i = 1, 2, \ldots, N$$

إن تقدير عدد العناصر في المجتمع يكون

$$\hat{N} = \sum_{i=1}^{n} 1/p_i = B\sum_{i=1}^{n} 1/w_i$$

وتقدير الكثافة للعناصر $D=N/A$ سيكون

$$\hat{D} = B\sum_{i=1}^{n}(1/w_i)/(BL) = (1/L)\sum_{i=1}^{n} 1/w_i$$

حيث إن n تمثل عدد العناصر التي قُطعت بخط التقاطع الذي طوله L. وتقدير الكثافة $D_y = \tau_y/A$ للمتغير y يمكن حسابه باستخدام

$$\hat{D}_y = (1/L)\sum_{i=1}^{n} y_i/w_i$$

أما تقدير المجموع الكلي $\tau = \sum_{i=1}^{N} y_i$ للمتغير y الذي تم قِيس في العناصر التي جرى قطعها من قبل خط القاطع فهو

$$\hat{\tau}_y = B\sum_{i=1}^{n} y_i/w_i$$

و تقدير الوسط الحسابي للمجتمع μ_y للمتغير y فيمكن الحصول عليه باستخدام

$$\hat{\mu}_y = (\sum_{i=1}^{n} y_i / w_i)(\sum_{i=1}^{n} 1 / w_i)$$

للحصول على تقدير لتباين التقديرات أعلاه لابد من سحب عينة عشوائية بحجم m من الخطوط. لذا يمكننا الحصول على تقديرات مستقلة لكثافة العناصر $(\hat{D}_1, \hat{D}_2, ..., \hat{D}_m)$ ولكثافة المتغير y، $(\hat{D}_{y_1}, \hat{D}_{y_2}, ..., \hat{D}_{y_m})$ في منطقة الدراسة. ومن ثم نجد الوسط الحسابي والتباين لهذه التقديرات لاستخدامها في التحليلات الإحصائية الأخرى مثل إيجاد فترة ثقة أو اختبار الفرضيات. هذا إذا كان أطوال الخطوط متساوية ولكن إذا كانت أطوال الخطوط غير متساوية فلابد من استخدام الوسط الحسابي المرجح لتقدير كثافة العناصر أو تقدير الكثافة للمتغير y التي ستكون على النحو الآتي:

$$\hat{D} = \sum_{k=1}^{m} L_k \hat{D}_k / \sum_{k=1}^{m} L_k$$

و

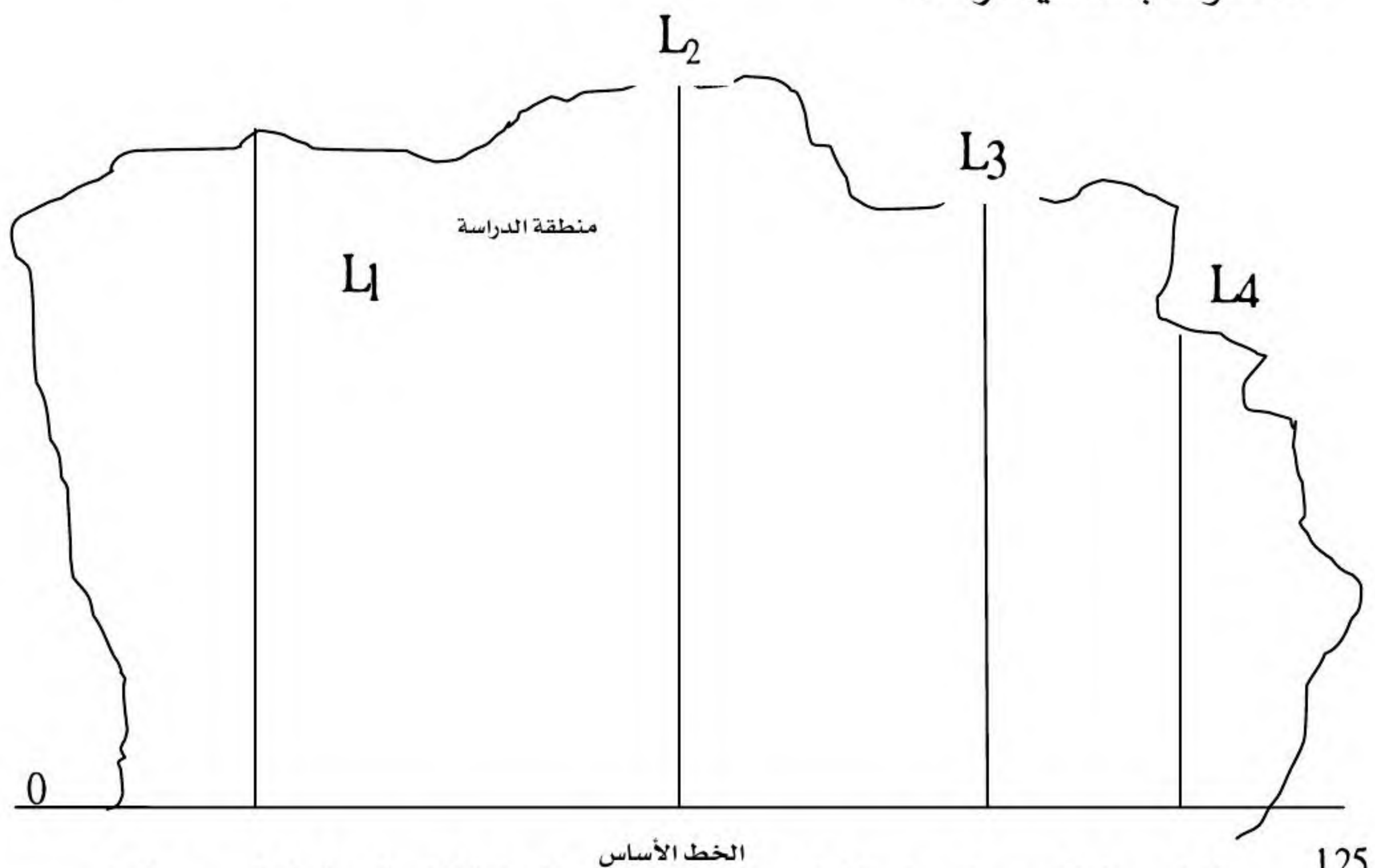
$$\hat{D}_y = \sum_{k=1}^{m} L_k \hat{D}_{y_k} / \sum_{k=1}^{m} L_k$$

إذا دققنا النظر في المعادلات الواردة في هذا الجزء سنجد أنها تعتمد على $w_i, i = 1, 2, ..., n$ الموجود في مقام هذه المعادلات. وبصورة عامة إذا كان حقل الدراسة كبير فإن احتمال أن يقطع العنصر i يكون $p_i = \frac{w_i}{B}$ حتى ولو كررنا الخطوط، لذا فإن قيمة صغيرة جداً إلى w_i سيؤدي إلى قيمة احتمال قريبة من الصفر وهذا سيؤدي إلى أن تكون التقديرات غير مستقرة

ولا يمكن الاعتماد عليها. ولابد في مثل هذه الحالات أن نحذف العناصر الصغيرة جداً التي تعطينا احتمالاً قريبًا من الصفر من العينة، وعلينا أن نَعُدَّها عناصر شاذة.

3.4.16 مثال

قام بهذه الدراسة مجموعة من الطلبة الباحثين لتقدير الكثافة والتغطية لإحدى الشجيرات المنتشرة في منطقة الدراسة، حيث قاموا بمد خط الأساس بطول 125 متر في منطقة الدراسة كما هو موضح في الشكل3 أدناه. ومن ثم قاموا بسحب أربعة أرقام عشوائية في الفترة (0.125) ومن موضع هذه النقاط قاموا بمد أربعة خطوط عمودية على الخط الأساس وقطعت منطقة الدراسة بأطوال مختلفة وهي على النحو الآتي $L_1 = 81$ و $L_2 = 106$ و $L_3 = 79$ و $L_4 = 50$ متر. الجدول 1 يعطينا v_i و w_i و y_i الارتفاع الأقصى للشجيرة لجميع الخطوط بالسنتيمترات.



الشكل3: عينة عشوائية من 4 خطوط ثابتة الاتجاه بأطوال مختلفة

الجدول 1: يعطينا الجدول v_i طول التقاطع بين الخط والعنصر (الشجيرة) و w_i ظل الشجيرة (عرض) على خط الأساس و y_i الارتفاع الأقصى للشجيرة للخطوط الأربعة.

الخط الأول L_1	v_{1i} = 268, 234, 193, 53, 50, 24, 13, 19, 43, 74, 72, 39, 40, 40 w_{1i} = 153, 87, 79, 78, 185, 145, 38, 52, 22, 38, 59, 20, 42, 102 y_{1i} = 170, 83, 110, 104, 141, 65, 28, 42, 24, 42, 31, 34, 29, 66
الخط الثاني L_2	v_{2i} = 52, 16, 95, 59, 112, 62, 44, 56, 86, 31, 15, 27, 53, 34, 38, 140, 142, 30 w_{2i} = 67, 31, 83, 195, 136, 145, 72, 115, 98, 129, 88, 26, 63, 112, 34, 21, 136, 95 y_{2i} = 67, 26, 70, 120, 125, 119, 58, 81, 39, 114, 78, 83, 39, 87, 6, 41, 113, 68
الخط الثالث L_3	v_{3i} = 190, 263, 60, 80, 40, 132, 94, 50, 150, 39, 118 w_{3i} = 115, 87, 57, 97, 57, 197, 58, 254, 185, 35, 124 y_{3i} = 120, 80, 40, 65, 50, 135, 93, 120, 95, 33, 96
الخط الرابع L_4	v_{4i} = 130, 175, 159, 182, 47, 4 w_{4i} = 96, 208, 68, 139, 50, 72 y_{4i} = 93, 143, 79, 86, 58, 63

استخدم البيانات لتقدير التغطية من قبل الشجيرات لهذه المنطقة، وكذلك قدر كثافة هذه الشجيرات في منطقة الدراسة، وأخيرًا كثافة المتغير y، الذي يمثل الارتفاع الأقصى للشجيرات في منطقة الدراسة. كذلك أوجد تقدير التباين والخطأ المعياري للتقديرات الثلاث على افتراض أن أطوال الخطوط متساوية.

الحل: لقد حُسبت التغطية لكل خط من الخطوط باستخدام العلاقة أعلاه، فلحساب التغطية للخط الأول على سبيل المثال استخدمنا

$$\hat{c}_1 = \sum_{i=1}^{14} v_i / L_1 = \frac{1}{81}(2.68+2.34+...+0.40) = 0.14346$$

الجدول 2 يعطينا التغطية لكل خط، أما التغطية لمنطقة الدراسة فقد استخدمنا الوسط الحسابي المرجح لحساب التغطية وهو

$$\hat{c} = \sum_{k=1}^{4} L_k \hat{c}_k / \sum_{k=1}^{4} L_k = [0.14346(81)+...+0.1394(50)]/(81+...+50) = 0.1319$$

أما فيما يخص حساب تقدير التباين والخطأ المعياري فقد افترضنا أن أطوال الخطوط متساوية واستخدمنا

$$s_c^2 = \frac{1}{m-1}\sum_{k=1}^{m}(\hat{c}_k - \bar{\hat{c}})^2 = \frac{1}{4-1}[(0.1435-0.1350)^2+...+(0.1394-0.1350)^2] = 0.00049284$$

حيث إن

$$\bar{\hat{c}} = \frac{1}{m}\sum_{k=1}^{m} \hat{c}_k$$

يمثل الوسط الحسابي غير المرجح، لذا فإن تقدير التباين والخطأ المعياري يكونا على التوالي:

$$s^2(\hat{c}) = \frac{s_c^2}{m} = \frac{0.00049294}{4} = 0.00012312$$

و

$$s(\hat{c}) = \sqrt{\frac{s_c^2}{m}} = \sqrt{0.00012312} = 0.0111$$

الآن نقوم بحساب الكثافة لكل خط من الخطوط الأربعة، فعلى سبيل المثال سنحسب الكثافة للخط الأول

$$\hat{D}_1 = 1/L_1 \sum_{i=1}^{14} 1/w_{1i} = [(1/1.53)+(1/0.87)+\ldots+(1/1.02)]/(81) = 0.33789$$

الجدول 2 يعطينا الكثافة لكل خط، أما الكثافة لمنطقة الدراسة فقد استخدمنا الوسط الحسابي المرجح لحسابها

$$\hat{D} = \sum_{k=1}^{4} L_k \hat{D}_k / \sum_{k=1}^{4} L_k = [81(0.337889)+\ldots+50(0.11425)]/(316) = 0.2384$$

ولحساب التباين والخطأ المعياري نفترض أن أطاول الخطوط متساوية ونستخدم العلاقة

$$s_{\hat{D}}^2 = \frac{1}{m-1} \sum_{k=1}^{m} (\hat{D}_k - \bar{\hat{D}})^2 = \frac{1}{4-1}[(0.1435-0.2235)^2+\ldots+(0.1394-0.2235)^2] = 0.01010$$

حيث إن

$$\bar{\hat{D}} = \frac{1}{m} \sum_{k=1}^{m} \hat{D}_k$$

يمثل الوسط الحسابي غير المرجح. لذا فإن تقدير التباين والخطأ المعياري يكونا على التوالي:

$$s^2(\hat{D}) = \frac{s_{\hat{D}}^2}{m} = \frac{0.01010}{4} = 0.002525$$

$$s(\hat{D}) = \sqrt{s^2(\hat{D})} = \sqrt{0.002525} = 0.0503$$

و أخيراً نقوم بحساب الكثافة للحد الأقصى لارتفاع الشجيرات للخطوط الأربعة، فعلى سبيل المثال سنحسبها للخط الأول

$$\hat{D}_{y_1} = 1/L_1 \sum_{i=1}^{14} y_{1i}/w_{1i} = [(1.7/1.53)+(0.83/0.87)+\ldots+(0.66/1.02)]/(81) = 0.16425$$

الجدول 2 يعطينا الكثافة للحد الأقصى لارتفاع الشجيرات لكل خط، أما الكثافة لمنطقة الدراسة فقد استخدمنا الوسط الحسابي المرجح لحسابها وهو:

$$\bar{D}_y = \sum_{k=1}^{4} L_k \bar{D}_{yk} / \sum_{k=1}^{4} L_k = [81(0.16425) + ... + 50(0.0919)]/(316) = 0.1395$$

ولحساب التباين والخطأ المعياري نفترض أن أطاول الخطوط متساوية ونستخدم العلاقة الآتية

$$s^2_{Dy} = \frac{1}{m-1}\sum_{k=1}^{m}(\bar{D}_{yk} - \bar{\bar{D}}_y)^2 = \frac{1}{4-1}[(0.16425 - 0.1332)^2 + ... + (0.091934 - 0.1.332)^2]$$

$$= 0.001225$$

حيث إن

$$\bar{\bar{D}}_y = \frac{1}{m}\sum_{k=1}^{m} \bar{D}_k$$

والذي يمثل الوسط الحسابي غير المرجح؛ لذا فإن تقدير التباين والخطأ المعياري يكونا على التوالي

$$s^2(\bar{D}_y) = \frac{s^2{}_{Dy}}{m} = \frac{0.001225}{4} = 0.0003063$$

$$s(\bar{D}_y) = \sqrt{s^2(\bar{D}_y)} = \sqrt{0.0003063} = 0.0175$$

جدول2: تقدير قيم التغطية والكثافة للارتفاع الأقصى لكل خط، وتقدير التغطية والكثافة للارتفاع الأقصى لمنطقة الدراسة باستخدام الوسط الحسابي المرجح والوسط الحسابي غير المرجح والخطأ المعياري للتقديرات المختلفة لمنطقة الدراسة.

	تقدير التغطية: $\hat{c}$	تقدير الكثافة: $\hat{D}$	تقدير الكثافة إلى: $\hat{D}_y$ y
الخط الأول L_1	0.14346	0.33789	0.16425
الخط الثاني L_2	0.10302	0.27235	0.16018
الخط الثالث L_3	0.15392	0.16947	0.11651
الخط الرابع L_4	0.1394	0.11425	0.09193
الوسط الحسابي المرجح	0.1319	0.2384	0.1395
الوسط الحسابي غير المرجح	0.1350	0.2235	0.1332
الخطأ المعياري	0.0111	0.0503	0.0175

References

1. Butler, S. A. and McDonald, L. L. (1980). Unbiased Systematic Sampling Plans for the Line Intercept Method, *Journal of Range Management*, 36, 463□468
2. Canfield, R. H. (1941). Application of the Line Intercept Method in Sampling Range Vegetation, *Journal of Forestry*, 39, 388□394.
3. DeVries, P. G. (1979). Line Transect Sampling□Statistical Theory, Applications and Suggestions foe Extended use Ecological Inventory. In R. M. Cormack, G. P. Patil and D. S. Robson (eds), Sampling Biological Population, Vol. 5 Statistical Ecology. Fairland, MD□ International Co□operative Publishing House, 1□70.
4. Eberhardt, L. L. (1978). Transect Methods for Population Studies. *Journal of Wildlife Management*. 42, 1□31.
5. Kaiser, L. (1983). Unbiased Estimation in Line□intercept Sampling. *Biometrics*, 39, 965□976.
6. Kimura, D. K. and Lemberg, N. A. (1981). Variability of Line Intercept Density Estimate□A Simulation Study of the Variance of Hydroacoustic Biomass Estimate. *Canadian Journal of Fisheries and Aquatic Sciences*, 38, 1141□1152.
7. Lucas, H. A. and Seber, G. F. (1997). Estimating Coverage and Particle Density using Line Intercept Sampling, *Biometrika*, 64, 618□622.
8. McDonald, L. L. (1980). Line□intercept Sampling for Attributes other than Coverage and Density. *Journal of Wildlife Management*, 44, 530□533.
9. Quzng, P. X. (1991). A Nonparametric Approach to Size□biased Line Transect Sampling, *Biometrics*, 39, 269□279.
10. Seber, G. A. F. (1979). Transect of random Length. In R. M. Cormack, G. P. Patil and D. S. Robson (eds), Sampling Biological Population, Vol. 5 Statistical Ecology. Fairland, MD□ International Co□operative Publishing House, 182□192.
11. Thompson, S. K (2002). Sampling, 2end Wiley, New York.

الفصل السابع عشر

المعاينة المكانية أو كريج انج
Spatial Sampling or Kriging

1.17 مقدمة

غالباً ما يكون مطلوباً في الدراسات الجيولوجية التنبؤ أو تقدير المخزون من المعادن أو النفط في موقع أو مكان معين، ويعتمد هذا التنبؤ على ما تمت مشاهدته في مواقع أو أماكن أخرى في المنطقة، وهذه المواقع غالباً ما تكون متباعدة عن بعضها بصورة غير منتظمة في المنطقة، وربما يكون مطلوباً التنبؤ أو تقدير المخزون من المعادن في المنطقة وليس فقط في موقع معين، فعند دراسة التلوث البيئي نقوم بقياس التلوث في موقع أو مواقع معينة، ومن ثم تستخدم هذه المواقع للتنبؤ أو تقدير مقدار التلوث البيئي في مواقع أخرى أو حتى منطقة أكبر. أخيراً يمكننا أن نستخدم المعاينة المكانية لتقدير أو التنبؤ بعدد الحيوانات أو النباتات الموجودة في منطقة الدراسة من خلال مشاهدة عينة من المواقع في هذه المنطقة.

لابد لنا أن نفكر في القيمة y_t من المعدن، أو النفط، أو التلوث، أو عدد الحيوانات في الموقع t بوصفها متغيرًا عشوائيًا، ونرغب في تقدير أو التنبؤ بقيمة متغير جديد y_0 باستخدام القيم $y_1, y_2, \dots, y_n$ التي شوهدت في n من المواقع $t_1, t_2, \dots, t_n$. والمتغير y_0 يمكن أن يكون قيمة المتغير في موقع جديد أو المجموع الكلي للمتغير في منطقة جغرافية أكبر. وبما أن y_0 يُعَدُّ متغيراً

عشوائياً وليس معلمة أو قيمة لصفة ثابتة للمجتمع، فإن الاستدلال الإحصائي يُعدُّ تنبؤاً وليس تقديراً بالرغم من أن التنبؤ يكون مكانياً وليس زمنياً.

إن مشكلة التنبؤ المكاني وحلها المسمى كريج انج (Kriging) يشبه إلى حد كبير طريقة التنبؤ أو التقدير باستخدام المعلومات الإضافية أو الخارجية كما مر معنا في الفصل السابع، ويمكن أن تكتب معادلات التنبؤ باستخدام التغاير (Covariance) أو باستخدام تباينات الفروق (Variogram)، علماً بأن التغاير يستخدم بصورة تقليدية مع السلاسل الزمنية، أما تباينات الفروق فتستخدم مع التنبؤ المكاني أو الإحصاء الجيولوجي (Geostatistics).

هناك مراجع كثيرة للتنبؤ المكاني، وكريج انج، والإحصاء الجيولوجي من أهمها: (1988) Hohn و(1986, 1989, 1991) Cressie و Journel (1987, 1988, 1989) وأيضًا هناك مراجع في سلسلة العمليات المكانية من أهمها Matern (1960, 1986) و(1983) Diggle و(1991)Cressie و Cox and Isham (1980) و(1981)Ripley.

2.17 التنبؤ المكاني

1.2.17 دالة التغاير المكانية (Spatial Covariance Function)

غالباً ما تكون قيمة المتغير الذي نرغب بدراسته غير مستقلة بين موقع وآخر في المسوحات البيئية والجيولوجية، بل -بالأحرى- تكون قيم المتغير في المواقع القريبة يعتمد بعضها على الآخر، ويمكننا أن نلخص العلاقة بين المتغيرين y_1 و y_2 المصاحبتين للموقعين t_1 و t_2 بالتغاير

$$\operatorname{cov}(y_1, y_2) = E[y_1 - E(y_1)][y_2 - E(y_2)]$$

عندما يكون التغاير بين موقعين يعتمد على موضعهما فقط، وليس على مكانهما المحدد في منطقة الدراسة، هذه العلاقة يمكن أن نلخصها بدالة

التغاير $c(h)$ عندما يكون التغاير للمتغير y لموقعين متباعدين عن بعضهما بمقدار h بما يأتي:

$$c(h) = cov(y_{t+h}, y_t)$$

أما إذا كانت منطقة الدراسة ثنائية الأبعاد فالموقع t سيعطى الإحداثيات بعدين أي $t = (t_1, t_2)$. كذلك المسافة h بين المواقع ستكون $h = (h_1, h_2)$، لذا فإن دالة التغاير ستعتمد على الاتجاه والمسافة بين المواقع.

2.2.17 التنبؤ الخطي (كريج انج) Linear Prediction (Kriging)

لنفترض أنه تم مشاهدة عينة بحجم n من المواقع وقياسها، ونرغب في تقدير ما سنجده في موقع آخر، وللتسهيل سنكتب y_i لقيمة المتغير y الذي تمت مشاهدته في الموقع i في العينة و$i=1, 2, \ldots, n$، وسنكتب c_{ij} بدلاً من $cov(y_i, y_j)$ للتغاير بين قيم y في الموقعين i وj. نلاحظ أن $c_{ii} = var(y_i)$. وباستخدام قيم المتغير y في المواقع $t_1, t_2, \ldots, t_n$ تجعلنا نرغب في التنبؤ بقيم المتغير y_0 في الموقع t_0. وللتسهيل أيضاً سنفترض أن الأوساط الحسابية $E(y_i)$ و$1=1, 2, \ldots, n$ متساوية.

إن الهدف الأساس هو إيجاد $\hat{y}_0$ التي هي عبارة دالة لقيم المتغير y والتي شوهدت باستخدام عينة بحجم n، والتي تكون تقديرًا غير متحيز إلى y_0، أي

$$E(\hat{y}_0) = E(y_0)$$

و كذلك تقوم بتصغير متوسط مربع خطأ التنبؤ أي

$$E(y_0 - \hat{y}_0)^2$$

يكون أصغر ما يمكن. بصورة عامة إن أفضل تقدير هو التوقع المشروط إلى y_0 إذا علمنا قيم $y_1, ..., y_n$ ، ولتحديد قيمة $E(y_0|y_1,...,y_n)$ لابد من معرفة التوزيع الاحتمالي المشترك إلى $y_1, ..., y_n$ وهذا سيكون صعباً.

بصورة عامة يمكن إيجاد دالة خطية لمشاهدات المتغير y تكون غير متحيزة، وتقوم بتصغير متوسط مربع خطأ التنبؤ، ويمكن كتابتها بالشكل الآتي

$$\hat{y}_0 = \sum_{i=1}^{n} a_i y_i$$

مشكلتنا الأساسية تكمن في كيفية إيجاد قيم $a_1, a_2, ..., a_n$ التي تقوم بتصغير متوسط مربع خطأ التنبؤ مع المحافظة على عدم التحيز. يمكننا أن نحصل على حل لهذه المشكلة باستخدام طريقة مضروب لكرا نج (Lagrange multiplier method) والتي يمكن كتابتها بصيغة المصفوفات

$$f = G^{-1} h$$

حيث إن

$$G = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1n} & 1 \\ c_{21} & c_{22} & \cdots & c_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ c_{n1} & c_{n2} & \cdots & c_{nn} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix} \text{ و } h = \begin{pmatrix} c_{10} \\ c_{20} \\ \vdots \\ c_{n0} \\ 1 \end{pmatrix} \text{ و } f = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \\ m \end{pmatrix}$$

علماً بأن الثابت m والذي تم الحصول عليه مع المعاملات a_i يمثل مضروب لكرا نج الذي يستخدم لإيجاد متوسط مربع خطأ التنبؤ.

إن أفضل مُتنبئ خطي في الجيولوجيا يدعى مُتنبئ كريج انج (Kriging predictor)، ومتوسط مربع خطأ التنبؤ إلى $\hat{y}_0$ يدعى تباين كريج انج وهو

$$E(y_0 - \hat{y}_0)^2 = c_{00} - \sum_{i=1}^{n} a_i c_{i0} - m$$

إذا كان المجتمع يتبع التوزيع الطبيعي، أي أن التوزيع الاحتمالي المشترك إلى $y_1, \ldots, y_n$ يتبع التوزيع الطبيعي المتعدد، فإن $\hat{y}_0 = \sum_{i=1}^{n} a_i y_i$ هو أفضل مُتنبئ ولديه أصغر متوسط لمربع الخطأ بالمقارنة بأي مُتنبئ آخر ولذا فهو دالة للمشاهدات $y_1, \ldots, y_n$.

إن الحصول على التنبؤ الأفضل يعتمد على معرفة دالة التغاير، أي أن تكون معلومة. ولكن في الحياة العملية دالة التغاير تكون غير معروفة ويجري تقديرها باستخدام البيانات التي تم الحصول عليها من المسح الحالي أو من المسوحات السابقة. ولكن إذا كانت السلسلة ثابتة ومتشابهة الخواص في جميع الاتجاهات، فإن التغاير للمواقع التي تبعد عن بعضها بمسافة d يمكن تقديرها باستخدام التغاير للعينة والتي تعتمد على n_d من الأزواج المتميزة للمواقع التي تبعد عن بعضها الآخر بمسافة مقدارها d تقريباً. لذا فإن تقدير التغاير للعينة يكون

$$\hat{c}(d) = \frac{1}{n_d} \sum (y_{ti} - \bar{y})(y_{tj} - \bar{y})$$

والمجموع يكون لجميع الأزواج المتميزة من المشاهدات التي تبعد عن بعضها الآخر بمسافة d وعدد هذه الأزواج هو n_d، ويمكننا أن نجد منحنىً توفيقياً ممهداً أو مهذباً (smooth curve) باستخدام طريقة المربعات الصغرى غير الخطية (nonlinear least squares) لتقدير دالة التغاير ومن ثم تقدير التغاير لأي مسافة.

مثال 1: أجريت بعض المسوحات لصيد الروبيان قرب إحدى الجزر القريبة من شواطئ ولاية ألاسكا الأمريكية لتقدير دالة التباين المكاني المشترك، التي سوف تستخدم للتنبؤ بكمية الروبيان التي يمكن صيدها في مواقع جديدة. لقد جرى رسم بيانات الصيد بالموضع على خريطة لمنطقة الدراسة، وأيضًا سُجلت كمية الصيد بالباون والمسافة بالميل البحري. وقد صُمم قارب البحث ليقوم برمي شبكة الصيد في مواضع تبعد عن بعضها مسافة ميل بحري، ومن ثم جرى حساب التغاير للعينة باستخدام أزواج من المشاهدات جمعت لفترات من المسافة. ومن ثم حسبنا المنحنى التوفيقي $a\ \exp(-bx)$ باستخدام طريقة المربعات الصغرى غير الخطية لتقديرات التغاير؛ لذا فإن الدالة التوفيقية للتباين المشترك ستكون

$$c(x) = 5.1\ e^{-0.49x}$$

لنفترض أنه تم رمي الشبكة مرة واحدة وصدنا $y_1 = 5.526$ ألف باوند من الروبيان، ومن ثم تم رمي الشبكة مرة ثانية على بعد 6 أميال بحرية وصدنا $y_2 = 1.417$ ألف باوند من الروبيان. ما هو تقديرك لكمية الروبيان y_0 الذي سنصيده إذا ألقينا الشبكة على بعد ميل بحري واحد من الرمية الأولى و5.4 أميالٍ من الرمية الثانية.

الحل: التباين سيكون $c(0)=5.1$، والتغاير لرميتين يبعدان عن بعضهما 6 أميال بحرية $c_{12} = 5.1\ e^{-0.49(6)} = 0.3$، التغاير للمواضع الجديدة: $c_{10} = 5.1\ e^{-0.49(1)} = 3.1$، و $c_{20} = 5.1\ e^{-0.49(5.4)} = 0.4$. لذا فإن دالة التنبؤ الخطي ستكون

$$\begin{pmatrix} a_1 \\ a_2 \\ m \end{pmatrix} = \begin{pmatrix} 5.1 & 0.3 & 1 \\ 0.3 & 5.1 & 1 \\ 1 & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} 3.1 \\ 0.4 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.78 \\ 0.22 \\ -0.95 \end{pmatrix}$$

لـذا فـإن $a_1 = 0.78$ ، و$a_2 = 0.22$ ، و$m = -0.95$. لـذا فـإن تنبؤنـا للصـيد في الموضع الجديد سيكون

$$\hat{y}_0 = 0.78(5.526) + 0.22(1.1417) = 4.662$$

أو 4662 باوند مقدار الصيد في الموضع الجديد. أما تقديرنا لمتوسط مربع خطأ التنبؤ فسيكون

$$\hat{E}(y_0 - \hat{y}_0)^2 = 5.1 - 0.78(3.1) - 0.22(0.4) + 0.95 = 3.5$$

وأما الخطأ المعياري لتقديرنا فسيكون $\sqrt{3.5} = 1.9$ أو 1900 باوند.

3.217 تباين الفروق (variogram)

غالباً ما يستخدم تباين الفروق (variogram) لوصف التغيرات المكانية في العلوم الجيولوجية بدلاً من التغاير، ويمكن تعريف تباين الفرق لقيم المتغير y في مواضع مختلفة كما يأتي:

$$\text{var}(y_{t+h} - y_t) = 2\gamma(h)$$

الدالة $\gamma(h)$ تسمى سيمافيريوجرام (semivariogram). عندما تكون العملية العشوائية مستقرة للدرجة الثانية فسيكون تباين الفروق والتغاير يحتوي على المعلومات نفسها، وذلك لأن

$$\gamma(h) = c(0) - c(h)$$

حيث $c(0) = \text{var}(y_t)$ وهو تباين y في الموقع t. لقد أشار Cressie(1986,1989,1991) إلى أن تباين الفروق يكون موجوداً لبعض العمليات العشوائية حتى وإن لم تكن مستقرةً للدرجة الثانية، لذا فإنها أكثر عمومية من التغاير.

لنفرض أن العملية كذلك مستقرة في الوسط الحسابي أي

$$E(y_t) = E(y_s)$$

لأي موقعين t و s في منطقة الدراسة. إذن

$$var(y_{t+h} - y_t) = E(y_{t+h} - y_t)^2$$

وبطريقة بسيطة يمكن أن نقدر $\gamma(h)$ باستخدام

$$2\hat{\gamma}(d) = \frac{1}{n_d}\sum(y_{t_i} - y_{t_j})^2$$

والمجموع يكون لجميع الأزواج المتميزة من المشاهدات التي تبعد عن بعضها بمسافة d وعدد هذه الأزواج هو n_d. للبيانات المتباعدة بشكل غير منتظم، يمكننا أن نجمع معاً الأزواج من المواقع التي تبعد عن بعضها تقريباً بنفس المسافة. وأخيراً يمكننا أن نجد منحنى توفيقياً ممهداً لتقدير تباينات الفروق لجميع المسافات. ولقد أشار Cressie(1991) إلى ضرورة تقدير واستخدام γ بدلاً من c دالة التغاير؛ لأن $\hat{\gamma}(d)$ تقدير غير متحيز إلى $\gamma(d)$، بينما $\hat{c}(d)$ تقدير متحيز إلى c(d).

يمكننا أن نكتب معادلات التنبؤ بدلالة السيمافيريوجرام. لنفترض أن المشاهدات تم قياسها في n من المواقع، ونرغب أن نتنبأ ماذا سنجد في موقع آخر. لتسهيل الرموز، سوف نستخدم الرمز y_i للدلالة على قيم المتغير y التي شوهدت في الموقع i في العينة، و$i=1,2,\ldots,n$، كذلك سنكتب λ_{ij} لترمز لقيمة السيمافيريوجرام $\gamma(y_i - y_j)$ للفرق بين الموقعين i وj. نرغب بالتنبؤ لقيمة المتغير y_0 في الموقع t_0، باستخدام قيم المتغير y في المواقع $t_1, t_2, \ldots, t_n$.

إن الهدف هو إيجاد الدالة $\hat{y}_0$ لمشاهدات المتغير y التي تكون تقدير غير متحيز إلى y_0 أي

$$E(\hat{y}_0) = E(y_0)$$

وتقوم بتصغير متوسط مربع خطأ التنبؤ

$$E(y_0 - \hat{y}_0)^2$$

ويمكننا كتابة المقدّر الخطي كما يأتي:

$$\hat{y}_0 = \sum_{i=1}^{n} a_i y_i$$

مشكلتنا كيف يمكننا أن نجد قيم $a_1, ..., a_n$ والتي تقوم بتصغير متوسط مربع خطأ التنبؤ مع المحافظة على عدم التحيز. الحل يمكن كتابة بصيغة المصفوفات

$$a = \Gamma^{-1} \gamma$$

حيث إنَّ

$$\Gamma = \begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1n} & 1 \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2n} & 1 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \gamma_{n1} & \gamma_{n2} & \cdots & \gamma_{nn} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{pmatrix} \text{ و } \gamma = \begin{pmatrix} \gamma_{10} \\ \gamma_{20} \\ \vdots \\ \gamma_{n0} \\ 1 \end{pmatrix} \text{ و } a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \\ m^* \end{pmatrix}$$

ومتوسط مربع خطأ التنبؤ إلى $\hat{y}_0$ هو

$$E(y_0 - \hat{y}_0)^2 = \sum_{i=1}^{n} a_i \gamma_{i0} + m^*$$

ملاحظة: $m^* = -m$ حيث إن m يمثل الثابت الذي وجدناه باستخدام دالة التغاير.

مثال 2: استخدم البيانات في المثال الأول لتقدير كمية الروبيان المتوقع صيده في الموقع الجديد باستخدام طريقة السيمافيريوجرام.

الحل: يمكننا إيجاد دالة السيمافيريوجرام باستخدام دالة التغاير التي وجدناه في المثال 1 وهي

$$\gamma(x) = 5.1\,(1 - e^{-0.49\,x})$$

لذا فإن قيمة السيمافيريوجرام لرميتين يبعدان عن بعضهما 6 أميال بحرية هي: $\gamma_{12} = 5.1\,(1 - e^{-0.49(6)}) = 4.8$. وعليه فإن قيمة السيمافيريوجرام للموقعين الجديدين ستكون: $\gamma_{10} = 5.1\,(1 - e^{-0.49(1)}) = 2.0$ و $\gamma_{20}=4.7$.

لذا فإن دالة التنبؤ الخطي باستخدام السيمافيريوجرام ستكون

$$\begin{pmatrix} a_1 \\ a_2 \\ m^* \end{pmatrix} = \begin{pmatrix} 0 & 4.8 & 1 \\ 4.8 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}^{-1} \begin{pmatrix} 2.0 \\ 4.7 \\ 1 \end{pmatrix} = \begin{pmatrix} 0.78 \\ 0.22 \\ 0.95 \end{pmatrix}$$

لذا فإن $a_1 = 0.78$ ، و $a_2 = 0.22$ ، و $m^* = 0.95$ ؛ لذا فإن تنبؤنا للصيد في الموضع الجديد سيكون

$$\hat{y}_0 = 0.78(5.526) + 0.22(1.1417) = 4.662$$

أو 4662 باوند مقدار الصيد قي الموضع الجديد ، أما تقديرنا لمتوسط مربع خطأ التنبؤ فسيكون

$$\hat{E}(y_0 - \hat{y}_0)^2 = 0.78(2.0) + 0.22(4.7) + 0.95 = 3.5$$

وأما الخطأ المعياري لتقديرنا فسيكون $\sqrt{3.5} = 1.9$ أو 1900 باوند.

للتنبؤ بقيمة المتغير y لمجتمع أو منطقة الدراسة وليس لموقع أو لمواقع محددة يراجع Thompson (2002).

3.17 التصميم المكاني

يعتمد متوسط مربع خطأ التنبؤ للمُتنبئ الخَطِّي غير المتحيز $\hat{y}_0$ لقيمة y_0 سواء كانت y_0 تمثل قيمة المتغير في موضع معين أو المتوسط لمنطقة الدراسة، على قيم التغاير بين y_0 وقيم العينة لكل موقع، كذلك على قيم التغاير بين المواقع المختلفة. لذا يمكننا أن نستخدم المعلومات المتوافرة من دالة التغاير أو تباينات الفروق لنقرر ما الذي نختاره من المواقع في العينة ليعطينا أفضل تنبؤ، بعبارة أخرى ما هو أفضل تصميم للمعاينة يمكن استخدامه للتنبؤ عن قيمة المتغير y_0.

لنفترض أن c_{ij} ترمز إلى دالة التغاير $\operatorname{cov}(y_i, y_j)$ بين قيم y في الموقعين i وj، ولنفترض أن c_{i0} تمثل التغاير بين y_i و y_0. بما أن $E(\hat{y}_0)=E(y_0)$، لذا فإن متوسط مربع خطأ التنبؤ إلى $\hat{y}_0 = \sum_{i=1}^{n} a_i y_i$ سيكون

$$E(y_0 - \hat{y}_0)^2 = \operatorname{var}(y_0 - \hat{y}_0) = \operatorname{var}(y_0) + \operatorname{var}(\hat{y}_0) - 2\operatorname{cov}(y_0, \hat{y}_0)$$

لذا

$$E(y_0 - \hat{y}_0)^2 = c_0 + \sum_{j=1}^{n}\sum_{i=1}^{n} a_i a_j c_{ij} - 2\sum_{i=1}^{n} a_i c_{i0}$$

إذا دققنا النظر بالمعادلة الأخيرة يتضح لدينا أن أفضل التنبؤات التي سنحصل عليها من عينة بحجم n من المواقع يكون عندما تكون قيمة التغاير قليلة بين المواقع وكبيرة بين المواقع و y_0. لنفترض أن دالة التغاير متماثلة بشكل دائري وتتناقص مع المسافة أو أن دالة تباين الفروق متماثلة وتتزايد مع المسافة. ويمكننا الحصول على قيمة كبيرة إلى c_{i0} وذلك عندما تكون المواقع المختارة بالعينة قريبة من الموضع 0 الذي نرغب في التنبؤ بقيمته. أما إذا كنا نرغب في التنبؤ عن قيمة الوسط الحسابي لمنطقة الدراسة فيفضل أن

تكون المواقع المختارة في العينة قريبة من مركز منطقة الدراسة. وللحصول على قيمة صغيرة إلى c_{ij} نقوم بتوزيع المواقع المختارة في العينة بصورة منتظمة على منطقة الدراسة أو بتقسيم منطقة الدراسة إلى عدد كبير من الطبقات الصغيرة.

للتنبؤ عن قيم y في مواضع منفردة في منطقة الدراسة، يمكننا أن نصغر قيمة أسو أخطاء التنبؤ باختيار المواقع التي عددها n لتصغير أقصى مسافة لأي نقطة في منطقة الدراسة من المواقع القريبة في العينة. لقد قام كل من McBratney et al.(1981 a,b) و Y fanits et al.(1987) بدراسة هذه المشكلة، وتم المقارنة بين شبكات مختلفة من نقاط العينة المنتظمة على شكل مربع، ومستطيل، ومثلث، ومسدس. ولقد تم التوصل إلى أن الشبكة المثلثة والمتساوية الأضلاع تؤدي إلى تصغير أبعد مسافة من نقاط العينة إلى النقاط الأخرى في منطقة الدراسة التي ليست موجودة في العينة.

قام Matern(1960,1986) بدراسة فاعلية مختلف تصاميم المعاينة للتنبؤ بقيمة الوسط الحسابي لمنطقة الدراسة. ولقد تم افتراض أن دالة التغاير متماثلة دائرياً ومتناقصة مع المسافة. لقد وجد للعينات المنتظمة أن الشبكة المثلثة أكثر فاعلية من الشبكة المربعة، بينما الشبكة المستطيلة وُجد بأنها أقل فاعلية من الشبكتين المثلثة والمربعة.

كذلك قام Matern بدراسة أفضل الطرق لتقسيم منطقة الدراسة إلى طبقات، وتوصل إلى أن أفضل التصاميم أو التصميم الأكثر فاعلية هو الذي يقوم بتقسيم منطقة الدراسة إلى طبقات صغيرة ومضغوطة الشكل، ولقد تم التوصل إلى أفضل شكل، أو الشكل الأكثر فاعلية للطبقة تكون على الترتيب الآتي: دائرة، ومسدس، ومربع، ومثلث متساوي الأضلاع. ولابد من الإشارة إلى أن معظم مناطق الدراسة لا يمكن تقسيمها إلى طبقات على

شكل دوائر. ولقد وجد أن الأشكال الثلاثة الأولى قريبة من بعضها من حيث الفاعلية. أما المثلث فهو أقل فاعلية من الأشكال الثلاثة الأولى. أما الطبقات المستطيلات فهي أقل فاعلية من المربعات، وتقل فاعليتها كلما ابتعدنا كثيراً من الشكل المربع.

إن أهم فوائد العينة المنتظمة تكمن في إمكانية نشر وحدات العينة بشكل متباعد إذا كانت الوحدات القريبة من بعضها مترابطة طردياً. ولكن العينة المنتظمة التي تكون وحداتها على شكل صف قد تكون غير فاعلة إذا كانت الوحدات المسحوبة بشكل منتظم (صف) تتوافق مع دورات متكررة في المجتمع، ويمكننا أن نسحب وحدات العينة المنتظمة بحيث لا تكون على شكل صف راجع Bellhouse(1988) كما يأتي: نقوم بتقسيم منطقة الدراسة المستطيلة الشكل إلى m_1m_2 من المربعات، يكون عدد الصفوف m_1 والأعمدة m_2. يكون طول ضلع المربع k. ومن ثم نقوم بسحب m_1 من الأعداد العشوائية من التوزيع المنتظم في الفترة $(0,k)$، والتي ستكون عبارة عن الإحداثيات للمحور الأفقي. ونقوم بعد ذلك بسحب m_2 من الأعداد العشوائية من التوزيع المنتظم في الفترة $(0,k)$، وهذه ستكون عبارة عن الإحداثيات للمحور العمودي، لذا فإن نقاط العينة للصف الأول والعمود الأول قد تم اختيارها، أما بقية نقاط العينة فيجري تحديدها بإضافة مضاعفات العدد الثابت k لكل نقطة من إحداثيات المحور الأفقي ولكل نقطة من إحداثيات المحور العمودي.

لقد أثبت Bellhouse(1981) أن العينة الطبقية التي تحتوي على وحدة واحدة في كل طبقة أكثر فاعلية من العينة المنتظمة التي جرى سحبها بالطريقة أعلاه. لقد قام Olea(1982,1984a,b) بدراسة استخدام تصاميم العينة الطبقية والمنتظمة واستخداماتهما في المعاينة الجيولوجية، كما قام

Barnes(1988) بالنظر إلى تحديد حجم العينة لتقدير القيم المتطرفة بدلاً من القيم المتوسطة في منطقة الدراسة، وقام Gilbert(1987) بدراسة وتلخيص بعض سمات المعاينة المكانية لمراقبة تلوث البيئة، وأخيراً قام McArthur(1987) بمقارنة مجموعة من تصاميم المعاينة والمقدّرات لتقدير متوسط تركيز التلوث في نقطة المصدر في منطقة الدراسة. ويُعَدُّ Muller(2001) أحدث مرجع للتصاميم المكانية.

References

1. Barnes, R. (1988). Bounding the Required Sample Size for Geologic Site Characterization, *Mathematical Geology*, 20, 447–490.
2. Bellhouse, D. R. (1981). Area Estimation by Point–Counting Techniques, *Biometrics*, 37, 303–312.
3. Bellhouse, D. R. (1988). Systematic Sampling, In P. R. Krishnaiah and C. R. Rao (eds.), Handbook of Statistics, Vol. 6, Sampling, Amsterdam, Elsevier Science Publisher, 125–145.
4. Cox, D. R. and Isham, V. (1980). Point Processes, Chapman & Hall, London.
5. Cerssie, N. (1986). Kriging Nonstationary Data, *Journal of the American Statistical Association*, 81, 625–634.
6. Cerssie, N. (1989). Geostatistics. *The American Statistician*, 43, 197–202.
7. Cerssie, N. (1991). Statistics for Spatial Data, Wiley, New York.
8. Diggle, P. J. (1983). Statistical Analysis of Spatial Point Patterns, Academic Press, New York.
9. Gilbert, R. O. (1987). Statistical Methods for Environmental Pollution Monitoring, Van Nostrand Reinhold, New York.
10. Hohn, M. E. (1988). Geostatistics and Petroleum Geology, Van Nostrand Reinhold, New York.
11. Isaaki, E. H. and Srivaslava, R. M. (1989). Introduction to Applied Geostatistics, Oxford University Press, Oxford.
12. Journel, A. G. (1987). Geostatistics for Environmental Sciences–An Introduction, Project CR 811893, Las Vegas, NV–U. S. Environmental Protection Agency, Environmental Monitoring Systems Laboratory.
13. Journel, A. G. (1988). Non–Parametric Geostatistics for Risk and Additional Sampling Assessment. In L. Keith (ed.), Principles of Environmental Sampling, Washington, DC–American Chemical Society, 45–72.
14. Journel, A. G. (1989). Short Course in Geology, Vol. 8, Fundamentals of Geostatistics in Five Lessons. Washington, DC–American Geophysical Union, 40 pp.
15. Matern, B. (1960). Spatial Variation. *Meddelanden fran Statens Skogsforskningsinstitut*, 45 (5).
16. Matern, B. (1986). Spatial Variation, 2nd ed., Springer–Verlag, Berlin.
17. McArthur, R. D. (1987). An Evaluation of Sample Design for Estimating a Locally Concentrated Pollutant, *Communications in Statistics – Simulation and Computation*, 16, 735–759.
18. McBratney, R. W., Webster, R. and Burgess, T. M. (1981a). The Design of Optimal Sampling Schemes for Local Estimation and Mapping of Regionalized Variables I, *Computer and Geostatistics*, 7, 331–334.
19. McBratney, R. W., Webster, R. and Burgess, T. M. (1981b). The Design of Optimal Sampling Schemes for Local Estimation and Mapping of Regionalized Variables II, *Computer and Geostatistics*, 7, 335–336.
20. Muller, W. G. (2001). Collecting Spatial Data–Optimum Design of Experiments for Random Fields, (2nd ed.), Physica–Verlag.

21. Olea, R. A. (1982). Optimization of the High Plains Aquifer Observation Network, Groundwater Series, No. 7 Lawrence, KS□Kansas Geological Survey, University Kansas, 73 pp.
22. Olea, R. A. (1984a). Sampling Design Optimization for Spatial Functions, *Mathematical Geology*, 16, 369□392.
23. Olea, R. A. (1982b). Systematic Sampling of Spatial Functions, Series on Spatial Analysis, No. 7 Lawrence, KS□Kansas Geological Survey, University Kansas, 57 pp.
24. Ripley, B. D. (1981). Spatial Statistics, Wiley, New York.
25. Thompson, S. K. (2002). Sampling, 2end Wiley, New York.
26. Yfantis, E. A., Flatman, G. T. and Behar, J. V. (1987). Efficiency of Kriging Estimation for Square, Triangular and Hexagonal Grids, *Mathematical Geology*, 19, 183□205

الفصل الثامن عشر

المعاينة العنقودية المكيفة
Adaptive Cluster sampling

1.18 مقدمة

المعاينـة المكيفـة (Apative sampling) تعـني أن الوحـدات الـتي سـيجري سـحبها لتكـون ضـمن العينـة ربمـا أنهـا سـتعتمد في عمليـة سـحبها علـى قيمـة المـتغير الـذي نرغـب بدراسـته وتمـت مشـاهدته خـلال المسـح. فمـثلاً إذا كـان المسـح مهتمـاً بتقـدير كثافـة نـوع مـن الحيوانـات أو النباتـات النـادرة في منطقـة الدراسـة، فسيجري ضم المواقع القريبة من الموقـع الـذي تم مشـاهدة الحيوانـات أو النبات النادر فيـه. كـذلك في دراسـة أنـواع معينـة مـن الأمـراض المعديـة فإنـه يجري ضم جميع أفراد العائلة إلى العينة إذا اكتشـف أن أحد أفرادهـا مصـاب بالمرض المعدي. تختلف المعاينة المكيفة عن طرق المعاينة التقليدية المعروفة مثل المعاينة العشوائية البسيطة أو غيرها، حيـث إن في هـذه الطـرق جميـع الوحـدات يجري تحديدها قبل البدء بعملية السحب للوحدات، ولا علاقة لقيمة المتغير الذي نرغب في دراسته بالوحدات التي سيجرى سحبها.

تسـتخدم العينـة العنقوديـة المكيفـة عنـدما يكـون المـتغير الـذي نرغـب في دراسـته علـى شـكل عناقيـد نـادرة الوجـود في منطقـة الدراسـة. لتنفيـذ المسـح باسـتخدام المعاينـة العنقوديـة المكيفـة، نقـوم بسـحب عينـة عشـوائية أوليـة مـن الوحـدات، وعنـدما تكـون قيمـة المـتغير الـذي نرغـب بدراسـته محققـة لـبعض

الشروط نقوم بسحب الوحدات المجاورة وإضافتها إلى العينة، فمثلاً عندما يقوم الباحثون بتقدير عدد الطيور أو الحيوانات النادرة أو كمية المعادن القلية التركيز أو عدد المصابين بنوع نادر ومعدٍ من الأمراض يقومون بسحب مواضع معينة في منطقة الدراسة أو المجتمع ومن ثم يقومون بزيارة هذه المواقع ولكنهم وفي معظم هذه المواقع المختارة لا يجدون أي أثر لما يبحثون عنه، ولكن حال عثورهم على وحدة تحتوي على أحد الطيور النادرة على سبيل المثال فإنه في الغالب ستكون الوحدات المجاورة تحتوي طيوراً أخرى لأن طبيعة بعض الحيوانات أو النباتات تعيش أو توجد في مناطق متجاورة؛ لذا لابد من إضافة الوحدات المجاورة إذا أردنا الحصول على تقدير جيد لما نريد تقديره. لذا تظهر لدينا أهمية استخدام العينة العنقودية المكيفة لتقدير معدل أو مجموع أو كثافة هذه الأنواع النادرة.

يعتبر Thompson(1990) أول من اقترح استخدام العينة العنقودية المكيفة على أن تستخدم العينة العشوائية البسيطة مع الإرجاع أو دون الإرجاع لسحب الوحدات الأولية، كما قام باقتراح استخدم العينة المنتظمة والطبقية لسحب الوحدات الأولية ببحثيه Thompson(1991 a,b). لقد قام كلٍ من Cormack(1988) وSeber(1986,1992) بمناقشة أهمية استخدام طرق المعاينة المكيفة للمعاينات الخاصة بالدراسات البيئية.

لقد شهدت العشر السنوات الأخيرة من القرن الماضي وبداية القرن الحالي تطورات كثيرة لاستخدام المعاينة المكيفة ومن بين أهم هذه التطورات: المعاينة العنقودية المكيفة بمرحلتين Salehi and Seber(1997b) والمعاينة العنقودية المزدوجة المكيفة لكلٍ من الباحثين Felix Medina and Thompson(1999) وFelixMedina(2000). كذلك قام كل من Pontius(1997) وRoesch(1993) وSmith et al.(1995) بدراسة العينة العنقودية المكيفة باحتمالات غير متساوية. لقد تمت دراسة المعاينة المكيفة للمربع

ألاتيني من قبل Munhollandand Brokowski(1993) و Borkowski(1999). كما تناول كل من Brown and Manly(1998) و Salehi and Seber(2002) المعاينة العنقودية المكيفة والمقيدة بحجم عينة محدد. هناك بعض النتائج التي من شأنها أن تسهل عملية الحسابات لنظرية Rao-Blackwell لتحسين المقدّرات التي تستخدم بالمعاينة المكيفة التي تمت دراستها من قبل Felix Medina(2003) و Salehi(1999). وهناك تطبيقات عديدة للمعاينة العنقودية المكيفة للمجتمعات الطبيعية والإنسانية والحيوانية وغيرها من المجتمعات المختلفة التي يهتم بها الباحثون لسبب أو لآخر تناولها العديد من الباحثين من بينهم Acharyal et al.(2000) و Boomer atel.(2000) و Clausen at el.(1999) و Petrucci(1998) و Danahera and King(1994). ولمزيد من المعلومات ولمواضيع متفرقة أخرى لاستخدامات أخرى للمعاينة العنقودية المكيفة يراجع: Thompson(1993,1996,2002) و Thompson et al.(1992) و Thompson and Seber(1996) و Muttlak and Khan(2002).

2.18 العينة العشوائية البسيطة كعينة أولية

1.2.18 سحب العينة العنقودية المكيفة

إنّ من أبسط طرق المعاينة العنقودية المكيفة تكون عندما نقوم بسحب الوحدات الأولية باستخدام العينة العشوائية البسيطة، حيث نقوم بتقسيم المجتمع إلى N من الوحدات الأولية ومن ثم نختار عينة عشوائية بسيطة بحجم n من الوحدات الأولية، ونقوم بإضافة الوحدات المجاورة عندما تحقق أياً من الوحدات الأولية شرطاً معيناً وليكن C، نقول: إن الوحدة i مستوفية للشرط إذا كانت قيمة المتغير الذي نرغب في دراسته يحقق الشرط أي $y_i \in C$. فعلى سبيل المثال الوحدة i مستوفية للشرط إذا كانت y_i أكبر أو تساوي رقماً ثابتاً c أي $C=\{y : y \geq c\}$. وإذا كان أيٌّ من الوحدات المضافة تحقق الشرط

نفسه نقوم بإضافة الوحدات المجاورة لها وهكذا إلى أن تكون جميع الوحدات الخارجية أو ما تسمى المجاورة لا ينطبق عليها الشرط C.

لشرح فكرة المعاينة العنقودية المكيفة لنفترض أننا نريد أن نقدر عدد حبات الكمأة أو ما يسمى في بعض المناطق الفقع لمنطقة الدراسة، وهي عبارة عن منطقة صحراوية مساحتها 150 كم^2. نقوم بتقسيم منطقة الدراسة إلى وحدات، وهي عبارة مربعات صغيرة مساحة كل وحدة 0.5 كم^2 كما هو مبين بالشكل 1. كما هو معروف فإن وجود الفقع يكون على شكل عناقيد ونادراً ما تجد حبة واحدة بمفردها. لنفترض أن الأرقام الموجودة داخل المربعات تمثل عدد حبات الفقع داخل كل مربع لتقدير عدد حبات الفقع في منطقة الدراسة نقوم بسحب عينة عشوائية بسيطة أولية بحجم n=10 من المربعات التي ميزناها بوضع علامة x في هذه المربعات. الآن نبدأ بالبحث في المربعات عن الفقع، فإذا عثرنا على فقع في أحد المربعات فإننا نقوم بالنظر في المربعات المجاورة لهذا المربع التي تكون إلى اليمين والشمال والأعلى والأسفل فإن عثرنا في أيٍّ من هذه المربعات المجاورة على فقع فإننا نقوم بإضافتها إلى العينة.

					14		6							
					17	13	12	x						
					12	15	18				7	16		
											12	18	6	
		X									x10	7		
		12	x4								5			
										2	3			
												x		
		15	10			x								
		8	11	13										
		10x	15	4						11	15			
		7	8						16	15	8			
						x				10	20			
													x	
	X													

ومن الشكل 1: المعاينة العنقودية المكيفة، لتقدير عدد حبات الفقع لمنطقة الدراسة التي حجمها 300 وحدة سحبنا عينة عشوائية بسيطة أولية بحجم 10 وحدات موضحة بالعلامة x.

ثم نقوم بفحص المربعات المجاورة للمربع الذي وجدنا فيه فقعاً ونضيفها إلى العينة، وهكذا إلى أن نصل إلى أن جميع المربعات الخارجية خالية من الفقع، عند ذلك نتوقف عن إضافة مربعات جديدة كما هو موضح بالشكل 2. وتكون هذه هي المعاينة العنقودية المكيفة والوحدات المضللة تمثل العينة العنقودية المكيفة.

على الرغم من كون العنقود يُعَدُّ مجموعة طبيعية، فإنه لا يمكن استخدام العناقيد أو العنقود للقيام بالاشتقاقات النظرية؛ وذلك للدور المزدوج الذي تلعبه الوحدات الموجودة على حافة العناقيد التي تسمى وحدات الحافة التي هي عبارة عن الوحدات الخالية كما هو موضح بالشكل 2، إذا تم اختيار إحدى وحدات الحافة بالعينة فهي تمثل عنقوداً بحجم 1، ولكن إذا لم يتم اختيار الوحدة في العينة الأولية فإنه لا يزال يمكن اختيارها كوحدة حافة في أحد العناقيد؛ لهذا سنقوم باستعمال فكرة الشبكة A_i للوحدة i وهي عبارة عن العنقود الذي تم الحصول عليه مع الوحدة i ولكن من دون وحدات الحافة، لذا فإن اختيار أي وحدة بالشبكة i سيؤدي إلى اختيار كل الشبكة A_i. إذا كانت i هي الوحدة الوحيدة التي ينطبق عليها الشرط $C=\{y: y\geq 1\}$ فسيكون لدينا شبكة بحجم 1. كذلك نعرف الوحدة التي جرى سحبها في العينة الأولية ولا ينطبق عليها الشرط C بالشبكة وبحجم 1. وهذا يعني أن جميع العناقيد وبحجم وحدة واحدة عبارة عن شبكات بحجم 1. كذلك أي وحدة من وحدات الحافة.

					14		6							
					17	13	12	x						
					12	15	18				7	16		
											12	18	6	
		x									x10	7		
		12	x4								5			
							X			2	3			
												x		
		15	10			x								
		8	11	13										
		x10	15	4						11	15			
		7	8						16	15	8			
						x				10	20			
													x	
	x													

الشكل2. المعاينة العنقودية المكيفة لتقدير عدد حبات الفقع لمنطقة الدراسة. المربعات المظللة التي عددها 54 مربعاً (وحدة) تمثل حجم العينة العنقودية المكيفة.

عبارة عن شبكة بحجم 1. لذا فإن أي عنقود يحتوي على أكثر من وحدة يمكن تقسيمه إلى شبكة أو شبكات بحجم 1 (شبكة لكل وحدة من وحدات الحافة). يمكننا أن نلاحظ أن العناقيد يمكن أن يكون بينها عناصر مشتركة وذلك من خلال وحدات الحافة، ولكن الشبكات ستكون منفصلة عن بعضها واتحادها يساوي عدد وحدات المجتمع.

2.2.1.8 تقدير الوسط الحسابي والمجموع الكلي للمجتمع

عندما يتم اختيار وحدات العينة الأولية دون إرجاع سيؤدي ذلك إلى أن تكون جميع الوحدات متميزة وذلك بسبب عدم الإرجاع، ولكن البيانات النهائية يمكن أن تحتوي على مشاهدات معادة وذلك بسبب أن بعض العناقيد ممكن أن تحتوي على أكثر من وحدة أولية؛ لذا ستكون الوحدة i في العينة

إما لأنها جزء من شبكة جرى سحب إحدى وحداتها بما فيها الوحدة i في العينة الأولية، أو لكونها إحدى وحدات الحافة لإحدى الشبكات لنفترض أن m_i تمثل عدد الوحدات في الشبكة التي تحتوي الوحدة i و a_i تمثل عدد الوحدات في الشبكات التي تكون الوحدة i وحدة حافة. إذا كانت الوحدة i ينطبق عليها الشرط C فإن $a_i = 0$ ، ولكن إذا كانت الوحدة i لا ينطبق عليها الشرط فإن $m_i = 1$. إن احتمال سحب الوحدة i في أي سحبة من السحبات المتعاقبة التي عددها n هو

$$p_i = \frac{m_i + a_i}{N}$$

أما احتمال أن تتضمن العينة الوحدة i فسيكون

$$\pi_i = 1 - \binom{N - m_i - a_i}{n} \Big/ \binom{N}{n}$$

عندما يجري سحب الوحدات الأولية مع الإرجاع، فسنحصل على مشاهدات معادة إما في العينة الأولية وذلك بسب عدم الإرجاع أو بسبب سحب أكثر من وحدة في العنقود. وباستخدام هذا التصميم سيكون احتمال سحب الوحدة i في أي سحبة من السحبات المتعاقبة هو

$$p_i = \frac{m_i + a_i}{N}$$

أما احتمال أن تتضمن العينة الوحدة i فسيكون

$$\pi_i = 1 - (1 - p_i)^n .$$

باستخدام أي التصميمين (دون إرجاع أو مع الإرجاع) فإنه لا يمكن حساب جميع الاحتمالات p_i أو π_i باستخدام المعلومات المتوافر في العينة وذلك بسبب عدم معرفة جميع قيم a_i.

لا يمكن استخدام المقدّرات المعتادة التي مرت معنا في المعاينة العشوائية البسيطة أو المعاينة العنقودية في المعاينة العنقودية المكيفة دون إدخال بعض التعديلات عليها؛ وذلك لكونها ستكون مقدّرات منحازة. ولكن إذا استخدمنا العينة الأولية على اعتبار أن الوحدات التي ينطبق عليها الشرط C تمثل عنقوداً بحجم 1 فإن الوسط الحسابي سيكون تقديراً غير متحيز لتقدير الوسط الحسابي للمجتمع، ولكنه أهمل جميع البيانات التي تمت إضافتها باستخدام المعاينة العنقودية المعدلة وعليه سيكون أقل فاعلية من المقدّرات التي سنتناولها في هذا الفصل.

سوف نكتفي بطريقة واحدة لتقدير الوسط الحسابي للمجتمع أو المجموع الكلي باستخدام المعاينة العنقودية المكيفة، وذلك عندما يتم سحب الوحدات بشكل متعاقب. ولمزيد من المعلومات يراجع Thompson(2002). لنفترض أن A_i تمثل الشبكة التي تحتوي على الوحدة i و m_i يمثل عدد الوحدات بالشبكة i والوحدة التي لا تستوفي الشرط يمكن اعتبارها شبكة بحجم 1. ولنفترض أن w_i يمثل متوسط المشاهدات للشبكة التي تحتوي على الوحدة i للعينة الأولية أي:

$$w_i = \frac{1}{m_i} \sum_{j \in A_i} y_j$$

لذا فإن المقدّر غير المتحيز للوسط الحسابي للمجتمع سيكون

$$\hat{\mu}_1 = \frac{1}{n} \sum_{i=1}^{n} w_i$$

أما التباين فسيكون

$$var(\hat{\mu}) = \frac{N-n}{n(N-1)} \sum_{i=1}^{N} \frac{(w_i - \mu)^2}{N}$$

إذا سحبت العينة الأولية من دون إرجاع، أما إذا سحبت العينة الأولية مع الإرجاع فسيكون التباين

$$var(\hat{\mu}) = \frac{1}{n} \sum_{i=1}^{N} \frac{(w_i - \mu)^2}{N}$$

أما تقدير التباين فسيكون

$$s^2(\hat{\mu}) = \frac{N-n}{nN} \sum_{i=1}^{n} \frac{(w_i - \hat{\mu})^2}{n-1}$$

إذا سحبت العينة الأولية من دون إرجاع، أما إذا سحبت العينة الأولية مع الإرجاع فسيكون التباين

$$s^2(\hat{\mu}) = \frac{1}{n} \sum_{i=1}^{n} \frac{(w_i - \hat{\mu})^2}{n-1}$$

مثال 1: العينة الموضحة بالشكل 1 و2 تم سحبها باستخدام العينة العشوائية البسيطة ومن دون إرجاع وبحجم $n=10$. قَدِّر عدد الفقع في منطقة الدراسة والخطأ المعياري للتقدير، علماً بأن الأرقام الموجودة داخل المربعات تمثل عدد الفقع في كل مربع أو وحدة.

الحل:

لتقدر الوسط الحسابي لعدد حبات الفقع في منطقة الدراسة والخطأ المعياري للتقدير لابد من حساب W_i

و $i=1,2,...,10$ و $w_1=\frac{76}{7}=10.857$ و $w_2=\frac{16}{2}=8$ و $w_3=\frac{101}{10}=10.1$

و $w_4=w_5=...=w_{10}=0$. وعليه سيكون تقدير الوسط الحسابي هو

$$\hat{\mu}=\frac{1}{10}(10.857+8+10.1+0+\ldots 0)=2.896$$

أما تقدير التباين إلى $\hat{\mu}$ فهو

$$s^2(\hat{\mu})=\frac{N-n}{Nn(n-1)}\sum_{i=1}^{n}(w_i-\hat{\mu})^2=\frac{300-10}{300(10)(10-1)}\{(10.857-2.896)^2+$$

$$+(8-2.896)^2+(10.1-2.896)^2+(0-2.896)^2+...+(0-2.896)^2=2.149$$

لذا فإن تقدير عدد الفقع في منطقة الدراسة (المجموع الكلي) سيكون

$$N\hat{\mu}=300(2.896)=869$$

وأخيراً فإن الخطأ المعياري للمجموع الكلي هو

$$\sqrt{N^2s^2(\hat{\mu})}=\sqrt{(300)^2(2.149)}=439.784$$

ولكن إذا استخدمنا الوسط الحسابي العادي للعينة النهائية فإننا سنحصل على $\bar{y}=193/54=3.574$: أما تقدير للمجموع الكلي للفقع فهو $N\bar{y}=300(3.574)=1072$. في الواقع إن الوسط الحسابي العادي سيؤدي إلى تقديرات أعلى مما يجب.

3.218 المقارنة بين المعاينة العنقودية المكيفة والمعاينة العشوائية البسيطة

لابد من الإشارة هنا إلى أن عدم التحيز للتصميم المكيف في هذا الفصل لا يعتمد على نوع المجتمع الذي نتعامل معه أو نرغب في دراسته؛ لأن عدم التحيز يعتمد على التصميم أو ما يسمى design-based؛ لذا فإن كون المعاينة

العنقودية المكيفة أكثر فاعلية من المعاينة العشوائية البسيطة، وهذا لا يعتمد على نوع المجتمع الذي نرغب في دراسته أو نسحب العينة منه.

لنفترض أن لدينا المعاينة العنقودية المكيفة مع عينة أولية بحجم n، جرى سحبها باستخدام المعاينة العشوائية البسيطة مع الإرجاع، وتم الحصول على المقدّر $\hat{\mu}$ لتقدير الوسط الحسابي للمجتمع μ. لنفترض أن تباين المجتمع المحدود

$$\sigma^2 = (N-1)^{-1}\sum_{i=1}^{N}(y_i-\mu)^2$$

لذا فإن تباين الوسط الحسابي للعينة العشوائية البسيطة بحجم عينة ثابت n^* هو

$$\sigma^2(N-n^*)/Nn^*$$

و بمقارنة هذا بتباين $\hat{\mu}$ $var(\hat{\mu})$ سنحصل على أن تباين المعاينة العنقودية المكيفة سيكون أصغر من تباين الوسط الحسابي للعينة العشوائية البسيطة بحجم n^* إذا كان

$$\left(\frac{1}{n}-\frac{1}{n^*}\right)\sigma^2 < \frac{N-n}{Nn(N-1)}\sum_{k=1}^{K}\sum_{i\in A_k}(y_i-w_i)^2$$

حيث إن k يمثل عدد الشبكات. نلاحظ أن الحد الأيمن من المتباينة يحتوي على التباين داخل الشبكة. لذا فإن المعاينة العنقودية المكيفة ستكون أفضل من المعاينة العشوائية البسيطة إذا كان التباين داخل الشبكة كبيراً. يراجع Thompson(2002) لمزيد من المعلومات.

3.18 العينة الطبقية كعينة أولية

بعد تقسيم المجتمع أو منطقة الدراسة إلى طبقات، نقوم بسحب عينة عشوائية طبقية أولية من المجتمع أو منطقة الدراسة، عندما تكون قيمة المتغير الذي نرغب في دراسته في أي وحدة من الوحدات التي جرى سحبها منطبقاً عليه الشرط نقوم بإضافة الوحدات المجاورة إلى العينة، كذلك نقوم بإضافة أي من الوحدات المجاورة للوحدات التي تم إضافتها حديثاً للعينة إذا تحقق الشرط وهكذا، إلى أن تكون جميع الوحدات الموجودة على الحافة لا ينطبق عليها الشرط، أو بعبارة أخرى خالية من المتغير الذي نرغب في دراسته.

يُعَدُّ تصميم المعاينة العنقودية الطبقية المكيفة من التصاميم المهمة من الناحية العملية؛ لأن كثيراً من المجتمعات تتوافر عنها معلومات يمكن أن تستخدم لتقسيم المجتمع إلى طبقات، ولكن توزيع التركيزات أو العناقيد في المجتمع قد لا يمكن التنبؤ به. نقوم في المعاينة الطبقية التقليدية بجمع الوحدات المتقاربة أو المشابهة لبعضها في طبقة بناءً على معلومات سابقة. أما في المعاينة العنقودية المكيفة فنقوم بالاستفادة من الحالة العنقودية التي تكون عليها الوحدات في المجتمع عندما يكون حجم وتوزيع العناقيد في المجتمع لا يمكن التنبؤ بها قبل بدء المسح بالعينة.

باستخدام المعاينة العنقودية الطبقية المكيفة تكون المقدّرات التقليدية (كالوسط الحسابي الطبقي للعينة) متحيزةً، لذا سنقترح مقدّرات تكون غير متحيزةً باستخدام هذا النوع من المعاينة. كذلك ستكون لدينا بعض التعقيدات مثل الاختيار في طبقة معينة قد يؤدي بسبب إضافة الوحدات للعينة إلى التجاوز إلى طبقة أخرى. أي سيكون لدينا في عينة واحدة وحدات من أكثر من طبقة وهذا يعني أن المشاهدات في الطبقات المختلفة ليست مستقلة كما هو معلوم عن المعاينة الطبقية التقليدية. ولقد جرى اقتراح مقدّرات معينة

لحل هذه الإشكالية. وأخيراً لابد من الإشارة إلى أن معظم ما هو موجود في هذا الجزء مأخوذ من Thompson(1991 b).

1.3.18 تصميم المعاينة العنقودية الطبقية المكيفة

لتنفيذ المعاينة العنقودية الطبقية المكيفة لابد أن يكون المجتمع مقسماً إلى L طبقة، وتحتوي الطبقة h و $h = 1,2,\ldots,L$ على N_h وحدة، ومجموع وحدات المجتمع نرمز لها N. ترمز u_{hi} إلى الوحدة i في الطبقة h ومع كل وحدة نعرف المتغير y_{hi} الذي نرغب في دراسته. لكل وحدة من وحدات المجتمع u_{hi} نعرف المجاور أو الجوار لهذه الوحدة بأنه عبارة عن مجموعة من الوحدات بما فيها الوحدة u_{hi} ينطبق على هذه الوحدات خاصية التماثل. التي تعني إذا كانت الوحدة $u_{h'i'}$ مجاورة للوحدة u_{hi} فإن الوحدة u_{hi} مجاورة للوحدة $u_{h'i'}$. ربما يكون المجاور لإحدى الوحدات يحتوي على وحدات من أكثر من طبقة. نقول إن الوحدة u_{hi} ينطبق عليها الشرط إذا كانت قيمة المتغير y لهذه الوحدة يقع في المجموعة المحددة C. على سبيل المثال يمكن أن تكون $C = \{ y \mid y \ge c \}$.

					14		6							
					x17	13	12							
					12	15	18				7	16		
											x12	18	6	
											10	7		
		x12	4								5			
										x2	3			
		x												
							x							
		15	10											
		8	11	13										
		10	15	4						11	15			
		7	8						16	15	8		X	
					x					10	20			
										x				
	x													

الشكل3: المعاينة العنقودية الطبقية المكيفة لتقدير عدد حبات الفقع لمنطقة الدراسة التي يوجد فيها طبقتان سحبنا عينة عشوائية بسيطة أولية بحجم 5 وحدات من كل طبقة وهي الموضح بالعلامة x.

نقوم بسحب عينة عشوائية أولية من مجتمع أو منطقة الدراسة باستخدام المعاينة العشوائية الطبقية، وهذا يعني أننا سنقوم بسحب عينة عشوائية بسيطة دون إرجاع من كل طبقة h من طبقات المجتمع وبحجم n_h ويكون السحب من كل طبقة بشكل مستقل عن الطبقات الأخرى. وعندما ينطبق الشرط على أي وحدة فإن الوحدات المجاورة لها يتم إضافتها إلى العينة إن لم تكن ضمن العينة الأولية، وأيضًا نقوم بإضافة أي من الوحدات المجاورة للوحدات التي جرى إضافتها إذا انطبق الشرط على الوحدات الجديدة، وهكذا حتى تكون العينة تحتوي على جميع الوحدات التي ينطبق عليها الشرط.

لتوضيح فكرة سحب الوحدات للعينة العنقودية الطبقية المكيفة لنفترض أن في مثالنا الأول حول الفقع الموضح في الشكل 1 أن منطقة الدراسة كانت مقسمة إلى طبقتين كما هو موضح في الشكل 3 أعلاه. وسحبنا عينة عشوائية بسيطة بحجم وحدات 5 من كل طبقة، والمربعات التي تحتوي على x تمثل العينة الطبقية الأولية التي جرى سحبها بصورة مستقلة من كل طبقة. تكون أي من الوحدات التي جري سحبها بالعينة الطبقية الأولية مستوفية للشرط إذا كان $C=\{y \mid y\geq 1\}$، أي إذا كانت أي من الوحدات تحتوي

على حبة فقع واحدة أو أكثر فإنها مستوفية للشرط. وكذلك جرى إضافة الوحدات

					14		6							
					x17	13	12							
					12	15	18				7	16		
											x12	18	6	
											10	7		
		x12	4								5			
										X2	3			
		x												
							x							
		15	10											
		8	11	13										
		10	15	4						11	15			
		7	8						16	15	8		X	
					x					10	20			
										x				
	x													

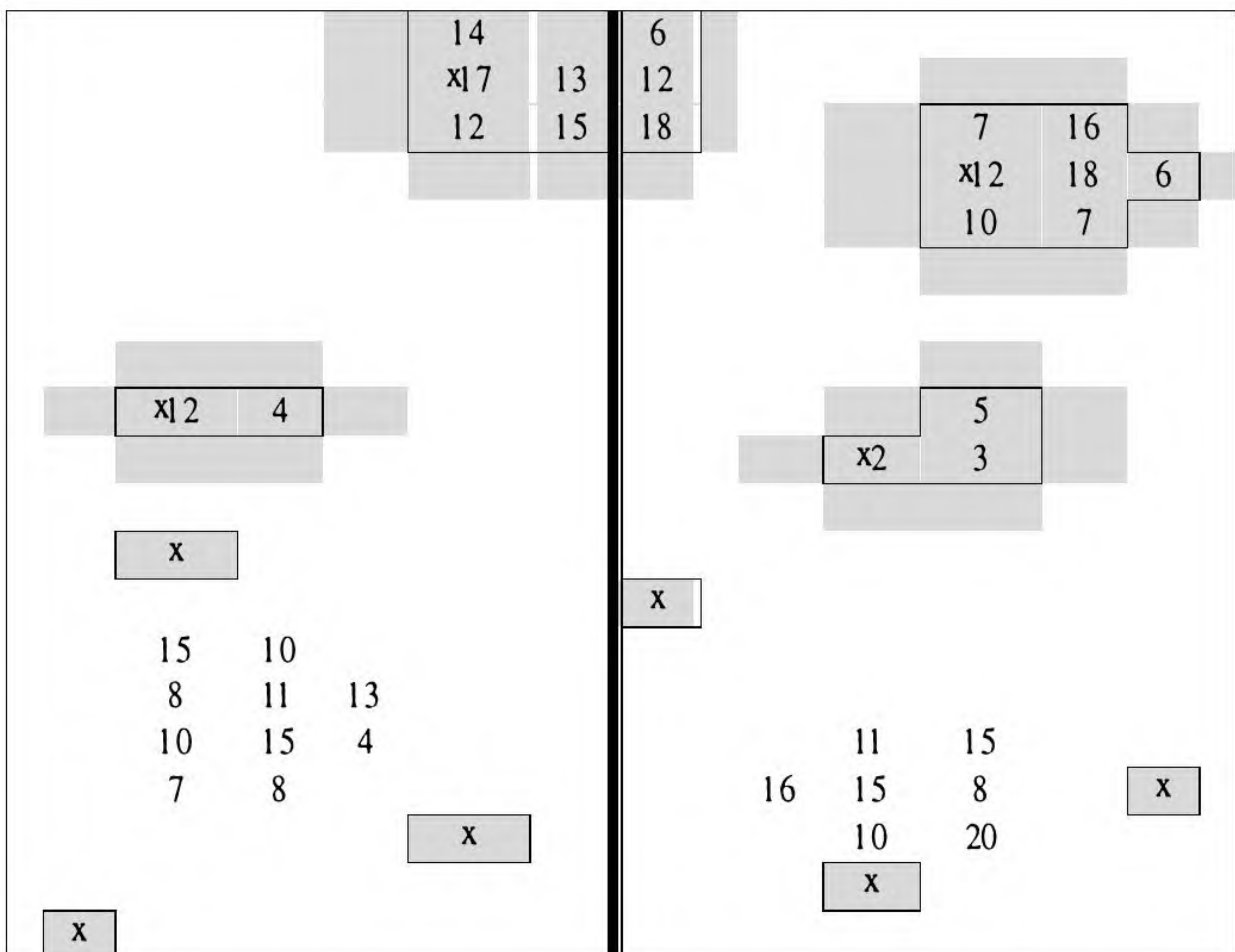
الشكل4: المربعات المظللة تمثل العينة العنقودية الطبقية المكيفة النهائية.
ملاحظة: بعض الوحدات من الطبقة الثانية في اليمين جرى إضافتها للعينة كنتيجة لاختيار إحدى وحدات الطبقة الأولى التي جرى سحبها في العينة الأولية من هذه الطبقة.

المجاورة للوحدات التي ينطبق عليها الشرط في المرحلة السابقة وهكذا. الشكل4 يوضح العينة العنقودية الطبقية المكيفة بشكلها النهائي.

يمكن تقسيم المجتمع إلى K من المجموعات ذات الوحدات والتي نسميها شبكات، بحيث إن اختيار أي وحدة في العينة الأولية سيؤدي إلى اختيار جميع الوحدات في الشبكة. الوحدة التي جرى سحبها بالعينة الأولية ولا تستوفي الشرط $C=\{y \mid y\geq 1\}$ هي عبارة عن شبكة بحجم 1. الشبكات المتميزة التي

تم الحصول عليها من خلال العينة الطبقية الأولية موضحة بالشكل 5 من خلال وضع خطوط حول وحدات كل شبكة.

لنفترض أن r_{hi} يمثل عدد المرات التي يمكن أن نختار بها الوحدة u_{hi}. ولنفترض أن m_{khi} يمثل عدد الوحدات التي تقع في تقاطع الطبقة k مع الشبكة التي تحتوي الوحدة u_{hi}. للوحدة u_{hi} والتي لا تستوفي الشرط ثم لنفترض أن a_{khi} تمثل مجموع الوحدات في تقاطع الطبقة k مع مجموعة من الشبكات المتميزة التي جرى تميزها بوضع خطوط حولها.



الشكل5: الشبكات المتميزة التي جرى الحصول عليها من العينة الأولية.

لا تحتوي الوحدة u_{hi} التي تتقاطع مع جيران الوحدة u_{hi}. الاختيار الأولي إلى أي من a_{khi} سيؤدي إلى إضافة الوحدة u_{hi} إلى العينة. لنعرف $a_{khi}=0$ لأي وحدة ينطبق عليها الشرط.

إن توقع عدد المرات التي نختار بها الوحدة u_{hi} سيكون

$$E(r_{hi})=\sum_{k=1}^{L} n_k \frac{m_{khi}+a_{khi}}{N_k}$$

23.18 تقدير الوسط الحسابي والمجموع الكلي للمجتمع

كما أشرنا سابقاً فإن المقدّرات التقليدية كالوسط الحسابي الطبقي سيكون متحيزاً في حالة استخدام المعاينة العنقودية الطبقية المكيفة، ويمكن أن نحصل على مقدّر غير متحيز لتقدير الوسط الحسابي للمجتمع ولو أنه غير فاعل وذلك باستخدام العينة العشوائية الطبقية، مع إهمال جميع الوحدات التي تمت إضافتها باستخدام المعاينة العنقودية الطبقية المكيفة.

للوحدة u_{hi} لنعرف متغير جديد نرمز له ب w_{hi}

$$w_{hi}=\frac{n_h}{N_h}\sum_{k=1}^{L}\xi_{khi} \Big/ \sum_{k=1}^{L}\frac{n_k}{N_k}m_{khi}$$

حيث إن ξ_{khi} يمثل مجموع قيم المتغير y الواقعة في الطبقة k والمتقاطعة مع الشبكة التي تحتوي على الوحدة u_{hi} و m_{khi} يمثل عدد الوحدات الواقعة في هذا التقاطع. لذا سيكون المقدّر للوسط الحسابي للمجتمع هو

$$\hat{\mu}_1=\frac{1}{N}\sum_{h=1}^{L}\frac{N_h}{n_h}\sum_{i=1}^{n_h}w_{hi}$$

أما التباين للمقدّر $\hat{\mu}_1$ فسيكون

$$var(\hat{\mu}_1)=\frac{1}{N^2}\sum_{h=1}^{L}N_h(N_h-n_h)\frac{\sigma_h^2}{n_h}$$

حيث إن

$$\sigma_h^2=\frac{1}{N_h-1}\sum_{i=1}^{N_h}(w_{hi}-\overline{W}_h)^2$$

و

$$\overline{W}_h=(1/N_h)\sum_{i=1}^{N_h}w_{hi}$$

للحصـول علـى تقـدير غـير متحيـز للتبـاين $var(\hat{\mu}_1)$ فـيمكن إيجـاده باستبدال σ_h^2 بتباين العينة s_h^2 لنحصل

$$s^2(\hat{\mu}_1)=\frac{1}{N^2}\sum_{h=1}^{L}N_h(N_h-n_h)\frac{s_h^2}{n_h}$$

حيث إن

$$s_h^2=\frac{1}{n_h-1}\sum_{i=1}^{n_h}(w_{hi}-\overline{w}_h)^2$$

و

$$\overline{w}_h=(1/n_h)\sum_{i=1}^{n_h}w_{hi}$$

أمـا تقـدير المجمـوع الكلـي للمجتمـع وتقـدير تباينـه فـيمكن إيجادهمـا باستخدام $N\hat{\mu}_1$ و $N^2s^2(\hat{\mu}_1)$ على التوالي.

يمكننا أن نحصل على مقدّر جديد $\hat{\mu}'$ لتقدير الوسط الحسابي للمجتمع وذلـك بتعـديل لمقـدّر $\hat{\mu}_1$ باتبـاع الطريقـة الـتي اقترحهـا كـل مـن

Birnbaum and Sirken(1965) و Levy(1977) و Sirken(1972) والتي يعتمد فيها الوزن على الطبقة التي توجد بها الوحدة التي سحبت في العينة الأولية وتتقاطع مع الشبكة. سنعرف للوحدة u_{hi} المتغير w'_{hi} الذي يمثل مجموع المتغير y في الشبكة التي تحتوي على الوحدة u_{hi} مقسوماً على عدد الوحدات في تلك الشبكة. لذا فإن

$$w'_{hi}=\sum_{k=1}^{L}\xi_{khi}/\sum_{k=1}^{L}m_{khi}$$

يمكننا أن نستخدم w'_{hi} للحصول على مقدّر غير متحيز للوسط الحسابي للمجتمع

$$\hat{\mu}'=\frac{1}{N}\sum_{h=1}^{L}\frac{N_h}{n_h}\sum_{i=1}^{n_h}w'_{hi}$$

والآن يمكننا أن نحصل على تباين للمقدّر $\hat{\mu}'$ وكذلك مقدّراً لتباينه باستخدام w'_{hi} بدلاً من w_{hi} في المعادلات الخاصة بالتباين وتقديره أعلاه.

مثال2: العينة العشوائية الطبقية والموضحة بالشكل 3 و4 و5 تم سحبها بعد تقسيم المجتمع إلى طبقتين حجمهما $N_1=140$ و$N_2=160$ ، حيث تم سحب خمس وحدات من كل طبقة كعينة أولية أي $n_1=5$ و$n_2=5$. قدر معدل عدد الفقع في منطقة الدراسة والخطأ المعياري للتقدير، علماً أنَّ الأرقام الموجودة داخل المربعات تمثل عدد الفقع في كل مربع أو وحدة.

الحل:

لنبدأ أولاً بإيجاد قيم w_{hi}

$$w_{11}=\frac{n_1}{N_1}\sum_{k=1}^{2}\xi_{khi}/\sum_{k=1}^{2}\frac{n_k}{N_k}m_{khi}=\frac{5}{140}(107)/[\frac{5}{140}(5)+\frac{5}{160}(3)]=14.033$$

$$w_{12}=\frac{n_1}{N_1}(\xi_{12})/\frac{n_1}{N_1}m_{12}=\frac{5}{140}(16)/\frac{5}{140}(2)=8$$

وبالطريقة نفسها نجد $w_{21}=76/7=10.857$ و $w_{22}=10/3=3.333$. نستطيع أن نحسب قيمة $\hat{\mu}$ لتكون

$$\hat{\mu}=(1/300)\{[140/5(14.033+8+0+0+0)]$$

$$+[160/5(10.857+3.33+0+0+0)]\}=3.57$$

لحساب تقدير التباين إلى $\hat{\mu}$ نريد أن نحسب $\bar{w}_h$ حيث $h=1,2$

$$\bar{w}_1=\frac{1}{n_h}\sum_{i=1}^{n_h} w_{hi}=\frac{1}{5}(17.125+8)=4.41$$

وبالطريقة نفسها نحسب $\bar{w}_2=2.838$. الآن نستطيع أن نحسب s_h^2 و $h=1,2$

$$s_1^2=[(14.033-4.41)^2+(8-4.41)^2+(0-4.41)^2+...+(0-4.41)^2=40.96$$

وبالطريقة نفسها نحسب $s_2^2=22.178$. الآن نستطيع أن نحسب تقدير التباين إلى $\hat{\mu}$

$$s^2(\hat{\mu})=\frac{1}{300^2}[140(140-5)\frac{40.96}{5}+160(160-5)\frac{22.178}{5}]=2.943$$

أما الخطأ المعياري للمقدّر فهو

$$\sqrt{s^2(\hat{\mu})}=\sqrt{2.943}=1.72 \, .$$

References

1. Acharyal, B., Bhattarai, G., de Gier, A. and Stein, A. (2000). Systematic Adaptive Cluster Sampling for the Assessment of Rare Tree Species in Nepal, *Forest Ecology and Management*, 137, 65–73.
2. Battista, T. D. (2003). Resampling Methods foe Estimating Dispersion Indices in Random and Adaptive Design, *Environmental and Ecological Statistics*, 10, 83–93.
3. Birnbaum, Z. W. and Sirken, M. G. (1965). Design of Sample Survey to Estimate the Prevalence of Rare Diseases: Three Unbiased Estimators, *Vital and Health Statistics*, Ser. 2, No. 11, Washington, DC: US Government Printing Office.
4. Boomer, K., Werner, C. and Brantley, S. (2000). CO_2 Estimation Related to the Yellowstone Volcanic System: 1. Developing a Stratified Adaptive Cluster Sampling Plan, *Journal of Geophysycal Research*, 105, 817–830.
5. Borkowski, J. J. (1999). Network Inclusion Probabilities and Horvitz–Thompson Estimation for Adaptive Simple Latin Square Sampling, *Environmental and Ecological Statistics*, 6, 291–311.
6. Brown, J. A. (2003). Design and Efficient Adaptive Cluster Sample, *Environmental and Ecological Statistics*, 10, 95–105
7. Brown, J. A. and Manly, B. F. J. (1998). Restricted Adaptive Cluster Sampling, *Environmental and Ecological Statistics*, 5, 47–62.
8. Christman, M. C. (2003). Adaptive Two–Stage One–Per–Stratum Sampling, *Environmental and Ecological Statistics*, 10, 43–60.
9. Clausen, D., Hanselman, D., Lunsford, C., Quinn, T. and Heifetz, J. (1999). Rockfish Adaptive Sampling Experiment in the Central Gulf of Alaska, 1998. AFSC Processed Report 99–04., Juneau, AK: Alaska Fisheries Science Center.
10. Cormack, R. M. (1988). Statistical challenges in the environmental sciences: A personal view, *Journal of the Royal Statistical Society*, Series A, 151, 201–210.
11. Danaher, P. J. and King, M. (1994). Estimating Rare Household Characteristics using Adaptive Sampling, *The New Zealand Statistician*, 29, 14–23.
12. Dryver, A. L. (2003). Performance of Adaptive Cluster Sampling Estimators in s Multivariate Setting, *Environmental and Ecological Statistics*, 10, 107–113.
13. Felix Medina, M.H. (2000). Analytical Expression for Rao–Blackwell Estimators in Adaptive Cluster Sampling, *Journal of Planning and Inference*, 84, 221–236.
14. Felix Medina, M.H. (2003). Asymptotics in Adaptive Cluster Sampling, *Environmental and Ecological Statistics*, 10, 61–82.
15. Felix Medina, M. H. and Thompson, S. K. (1999). Adaptive Cluster Double Sampling, Proceeding of the Survey Research Section, American Statistical Association, 86, 445–449.
16. Levy, P. S. (1977). Optimum Allocation in Stratified Random Network Sampling for Estimating the Prevalence Attributes in Rare Populations, *journal of the American Statistical Association*, 72, 758–763.
17. Munholland, P. L. and Borkowski, J. J. (1993). Adaptive Latin Square Sampling +1 Designs, Technical Report No. 3–23–93, Department of Mathematical Sciences, Montana State University, Bozeman.
18. Muttlak, H. A and Khan, A., (2002). Adjusted Two–sage Adaptive Cluster Sampling, *Environmental and Ecological Statistics*, 9, 111–120.
19. Petrucci, A. (1998). Adaptive Sampling for Environmental Pollution Data: Some Simulation Results, *Statistica Applicata*, 103.
20. Pontius, J. S. (1997). Strip Adaptive Cluster Sampling: Probability Proportion to Size Selection of Primary Units, *Biometrics*, 53, 1092–1096.
21. Roesch, F. A., Jr. (1993). Adaptive Cluster Sampling for Forest Inventories, *Forest Science*, 39, 655–669.

22. Salehi, M. M. (1999). Rao-Blackwell Versions of the Hansen-Hurwitz and Horvitz-Thompson Estimators in Adaptive Cluster Sampling, *Environmental and Ecological Statistics*, 6, 183-195.
23. Salehi, M. M. and Seber, G. F. A. (1997a). Adaptive Cluster Sampling with Networks Selected without Replacement, *Biometrika*, 84, 209-219.
24. Salehi, M. M. and Seber, G. F. A. (1997b). Two-Stage Adaptive Cluster Sampling, *Biometrics*, 53, 959-970.
25. Salehi, M. M. and Seber, G. F. A. (2002). Unbiased Estimators for Restricted Adaptive Cluster Sampling, *Australian and New Zealand Journal of Statistics*, 44, 63-75.
26. Salehi, M. M. (2003). Comparison Between Hansen-Hurwitz and Horvitz-Thompson Estimators for Adaptive Cluster Sampling, *Environmental and Ecological Statistics*, 10, 115-127.
27. Seber, G. F. A. (1986). A Review of Estimating Animal Abundance, *Biometrics*, 42, 267-292.
28. Seber, G. F. A. (1992). A Review of Estimating Animal Abundance- II *International Statistical Review*, 60, 129-166.
29. Sirken, M. G. (1972). Stratified Sample Survey with Multiplicity, *journal of the American Statistical Association*, 67, 257-266.
30. Smith, D. R., Conroy, M. J., and Brakhage, D. H. (1995). Efficiency of Adaptive Cluster Sampling for Estimating Density of Wintering Waterfowl, *Biometrics*, 51, 777-778.
31. Smith, D. R., Villella, R. F. and Lemarie, D. P. (2003). Application of Adaptive Cluster Sampling to Low-Density Populations of Freshwater Mussels, *Environmental and Ecological Statistics*, 10, 7-15.
32. Su Z. and Quinn, T. J. (2003). Estimators Bias and Efficiency for Adaptive Cluster Sampling with Order Statistics and a Stopping Rule, *Environmental and Ecological Statistics*, 10, 17-41.
33. Thompson, S. K. (1990). Adapive Cluster Sampling, *journal of the American Statistical Association*, 85, 1050-1059.
34. Thompson, S. K. (1991a). Adaptive Cluster Sampling-Design with Primary and Secondary units, *Biometrics*, 47, 1103-1115.
35. Thompson, S. K. (1991b). Stratified Adaptive Cluster Sampling, *Biometrika*, 78, 378-397.
36. Thompson, S. K. (1996). Adaptive Cluster Sampling based on Order Statistics, *Enviornmetrics*, 7, 123-133.
37. Thompson, S. K (2002). Sampling, 2end Wiley, New York.
38. Thompson, S. K., Ramsey, F. L. and Seber A. F. G. (1992). An Adaptive Procedure for Sampling Animal Populations, *Biometrics*, 48, 1195-1199.
39. Thompson, S. K. and Seber A. F. G. (1996). *Adaptive Sampling*, Wiley-New York.

الفصل التاسع عشر

المعاينة الشبكية
Network sampling

1.19 مقدمة

لتقدير انتشار أحد الأمراض النادرة، نقوم بسحب عينة عشوائية من المستشفيات أو المراكز الطبية في مجتمع أو منطقة الدراسة، ونقوم بفحص سجلات هذه المستشفيات للبحث عن المرضى الذين عولجوا في هذه المستشفيات بسبب هذا المرض النادر، ونستخرج سجلاتهم ونقوم بدراستها أو جمع معلومات منها. ولكن بعض المرضى لديهم سجلات في أكثر من مستشفى أو مركز طبي لنفس المرض أي عولجوا في أكثر من مستشفى. من الواضح أنه كلما عولج أحد المرضى في أكثر من مستشفى فإن احتمال أن يسحب سجله في العينة يكون أكبر.

في دراسة أخرى لنفرض أننا نريد أن نقدر انتشار بعض الصفات النادرة في المجتمع، لتحقيق ذلك نقوم بسحب عينة عشوائية بسيطة من بعض المنازل في مجتمع الدراسة، ونسأل ساكنيها البالغين عن حدوث هذه الصفة التي نرغب بدراستها ولكن ليس فقط عن حدوثها معهم بل مع جميع أشقائهم أو شقيقاتهم. لذا من الواضح أن الشخص الذي يسكن في منزل آخر ولديه عدد كبير من الأشقاء والشقيقات سيكون احتمال سحبه في العينة أكبر من الشخص الذي ليس لديه أشقاء يسكنون في منازل مستقلة، حتى في البيت

الواحد ليس من الضرورة أن تكون احتمالات ضم جميع الأشخاص البالغين والساكنين في البيت نفسه متساوية.

يطلق على هذا التصميم المعاينة الشبكية أو المعاينة المتعددة. في هذا النوع من المعاينة نقوم بسحب عينة عشوائية بسيطة أو عينة عشوائية طبقية من الوحدات تسمى الوحدات المختارة، ومن ثم نقوم بإضافة جميع الوحدات المشاهدة التي لها علاقة بالوحدات المختارة والتي سحبت بالعينة. على سبيل المثال إذا قمنا بسحب عينة عشوائية بسيطة من المنازل في مدينة ما وقمنا بسؤال الأشخاص البالغين والساكنين في هذه المنازل عن صفة معينة نادرة يتصفون بها هم وأشقاؤهم الساكنون في المدينة، وليس بالضرورة في المنزل نفسه الذي يسكنون فيه، فالمنازل في هذه الدراسة تمثل الوحدات المختارة وأما الأشخاص فيمثلون وحدات المشاهدة. تعرف تعددية (Multip licity) الشخص بأنها عبارة عن عدد الوحدات المختارة (مستشفى أو منزل كما في أمثلتنا أعلاه) وهي التي يكون للشخص علاقة بها، ونعرف الشبكة على أنها عبارة عن مجموعة من الوحدات التي تمت مشاهدتها والتي ترتبط ببعض ها الآخر برباط معين. يمكن أن ترتبط الشبكة بأكثر من وحدة مختارة (أشقاء يعيشون في أكثر من منزل). كذلك فإن أي وحدة مختارة يمكن أن ترتبط بأكثر من شبكة (أشخاص بالغين يعيشون في بيت واحد وليسوا بأشقاء). إذا كان المجتمع يحتوي على طبقات يمكن أن تتقاطع الشبكة مع أكثر من طبقة.

بما أن الوحدات يجري سحبها باحتمالات غير متساوية، فإن الوسط الحسابي للعينة سيكون مقدّراً متحيزاً للوسط الحسابي للمجتمع باستخدام هذا النوع من المعاينة. قام كلٌّ من Birnbaum and Sirken (1965) باقتراح مجموعة من المقدّرات غير المتحيزة. هنالك مجموعة من البحوث ركزت على المقدّر المتعدد (Multiplicity Estimator) قام بها الباحثون: Nathan(1976)

و(Sirken(1970,1972a,b أما (Levy(1977 و Sirken and Levy وقد ركزوا على نسب المقدّرات المتعددة التي يمكن أن تستعمل لتقدير نسبة مجموعة عرقية مصابة بنوع نادر من الأمراض. بينما ركز(Czaja et al.(1986 على تأثير الأخطاء الناتجة من جمع المعلومات عن طريق الترابط بين الشبكات، على سبيل المثال يمكن الاعتماد على المريض بصورة أكبر من الاعتماد على قريبه في المنزل للحصول على المعلومات المطلوبة. هنالك تطبيقات كثيرة ومتعددة للعينة الشبكية. يراجع كلٌّ من: (Kalton and Anderson(1986 و(Faulkenberry and Garoui(1991 و(Sudman et al.(1988.

2.19 تقدير الوسط الحسابي والمجموع الكلي

لنفترض أن قيمة المتغير الذي نرغب في دراسته للوحدة i في المجتمع هو y_i. في المسوحات التي نرغب في تقدير عدد المرضى المصابين بنوع نادر من الأمراض تكون قيمة $y_i = 1$ إذا كانت الوحدة تحمل الصفة، وهي أن يكون المريض مصاباً بالمرض النادر، وخلاف ذلك تكون قيمة المتغير $y_i = 0$. يمكن أن يأخذ المتغير y_i أي قيمة وليس فقط 1 و0. فعلى سبيل المثال يمكن أن تكون قيمة تمثل تكلفة العلاج للمريض i المصاب بالمرض النادر. لنفترض أن N تمثل عدد وحدات المجتمع التي يمكن مشاهدتها. أما المجموع الكلي للمجتمع فيعرف

$$\tau = \sum_{i=1}^{N} y_i$$

لنفترض أن m_i تمثل تعددية الوحدة المشاهدة i، أي عدد الوحدات المختارة التي ترتبط بوحدة المشاهدة. سوف نرمز لعدد الوحدات المختارة من المجتمع بـM. لذا فإن الوسط الحسابي للمجتمع للوحدات المختارة سيكون

$$\tau = t/M$$

سوف نستخدم العينة العشوائية البسيطة دون إرجاع لاختيار n من الوحدات ذالمختارة وكل وحدة مشاهدة ترتبط بأي وحدة من الوحدات المختارة سيجري ضمها إلى العينة.

سوف نستخدم المقدّر المتعدد لتقدير مجموع أو الوسط الحسابي للمجتمع. ولمزيد من المعلومات ولمقدّرات أخرى يراجع Birnbaum and Sirken(1965) أو Thompson(2002) إن احتمال سحب وحدة المشاهدة i في أي سحبة من السحبات المتعاقبة هو

$$p_i = \frac{m_i}{M}$$

يمكننا أن نجد المقدّر غير المتحيز للمجموع الكلي للمجتمع τ بقسمة قيمة المشاهدة للمتغير y على احتمال اختيار الوحدة المصاحبة، لذا فإن المقدّر المتعدد للمجموع الكلي يمكن تعريفه بما يأتي:

$$\hat{t}_m = \frac{1}{n}\sum_{i\in s}\frac{y_i}{p_i} = \frac{M}{n}\sum_{i\in s}\frac{y_i}{m_i}$$

حيث إن s تمثل وحدات المشاهدة المتعاقبة في العينة بما فيها الوحدات المعادة. لذا فإن وحدات المشاهدة يمكن سحبها أكثر من مرة بالرغم من أننا نسحب الوحدات دون إرجاع؛ لأن وحدات المشاهدة يمكن أن ترتبط بأكثر من وحدة من الوحدات المختارة؛ لذا فإن توقع عدد المرات التي يمكن أن تختار بها وحدة المشاهدة i هو np_i.

يمكننا أن نعيد تعريف المقدّر المتعدد للمجموع الكلي للمجتمع بشكل أكثر سهولة، للوحدة المختارة j في المجتمع نعرف المتغير w_j ليكون مجموع y_i/m_i لجميع وحدات المشاهدة المرتبطة بالوحدة j أي

$$w_j = \sum_{i\in A_j}\frac{y_i}{m_i}$$

حيث إن A_j عبارة عن وحدات المشاهدة المرتبطة بالوحدة المختارة j. وباستخدام الرموز الجديدة يمكننا أن نعيد كتابة المقدّر المتعدد للمجموع الكلي كما يأتي:

$$\hat{\tau}_m = \frac{M}{n}\sum_{j=1}^{n} w_j$$

يمكننا أن نفكر بالمتغير w_j بصفته متغيرًا جديدًا متعلقًا بالوحدة المختارة j ونرغب في دراسته. لذا سيكون المقدّر المتعدد عبارة عن $M\bar{w}$ ، حيث إن $\bar{w} = \frac{1}{n}\sum_{j=1}^{n} w_j$ عبارة عن الوسط الحسابي للعينة العشوائية البسيطة وبحجم n. لذا وباستخدام النتائج الأساسية للعينة العشوائية البسيطة نحصل على تباين $\hat{\tau}_m$ كما يأتي:

$$\operatorname{var}(\hat{\tau}_m) = \frac{M(M-n)}{n}\sigma_w^2$$

حيث إن

$$\sigma_w^2 = \frac{1}{M-1}\sum_{j=1}^{M}(w_j-\mu)^2$$

حيث إن $m = \tau / M$ تمثل الوسط الحسابي للمجتمع لكل وحدة اختيار. أما المقدّر غير المتحيز لتباين المجتمع $\operatorname{var}(\hat{\tau}_m)$ فسيكون

$$s^2(\hat{\tau}_m) = \frac{M(M-n)}{n}s_w^2$$

حيث إن

$$s_w^2 = \frac{1}{n-1}\sum_{j=1}^{n}(w_j-\bar{w})^2$$

لتقدير الوسط الحسابي لمجتمع الوحدة المختارة فسيكون

$$\hat{\mu}_m = \hat{\tau}_m / M$$

و تباين

$$\text{var}(\hat{\mu}_m) = \text{var}(\hat{t}_m) / M^2$$

و تقدير للتباين

$$s^2(\hat{\mu}_m) = s^2(\hat{t}_m) / M^2$$

مثال:

لشرح طريقة حسابات مقدّرات العينة الشبكية سوف نستخدم المثال الآتي: لتقدير شيوع أو انتشار أحد الأمراض في إحدى المدن والمؤلفة من M=5000 منزل، سحبنا عينة عشوائية بسيطة وبحجم n=100 منزل. البالغون في المنازل المختارة سيقومون بالتبليغ فيما إذا كانوا مصابين بالمرض أو أياً من أشقائهم أو شقيقاتهم في المدينة، هنا المنازل تمثل وحدات الاختيار أو الوحدات المختارة بينما الناس البالغون يمثلون وحدات المشاهدة، المتغير الذي نرغب في دراسته $y_i = 1$ إذا كان الشخص مصاباً بالمرض و $y_i = 0$ إذا كان الشخص غير مصاب بالمرض.

نقوم بترتيب المنازل والبالغ عددها 100 في العينة لنضع المنازل التي تحتوي على إصابات في المقدمة أي المنازل التي تكون فيها قيمة المتغير y لا تساوي صفراً بالمقدمة. يعيش في البيت الأول شخصان بالغان رجل وامرأة، لدى الرجل شقيق يسكن في دار منفصل، الرجل لا يحمل المرض $(y_1 = 0)$ ولكن شقيقه مصابٌ $(y_2 = 1)$، هؤلاء الشقيقان يشكلان الشبكة الأولى بتعددية تساوي 2 أي $m_1 = 2$، أما المرأة في البيت الأول فهي مصابة بالمرض $(y_3 = 1)$ ولديها شقيقان يسكنان في منزل مستقل، الأول مصاب بالمرض $(y_4 = 1)$ والثاني غير مصاب بالمرض $(y_5 = 0)$. هؤلاء الأشقاء الثلاثة يشكلون الشبكة الثانية بتعددية مقدارها 3 أي $m_2 = 3$، ولكن المنزل الذي يسكن فيه الشقيق الخامس جرى سحبه في العينة أي من بين المئة منزل، لذا فإن المنزلين الذين يسكن فيهما الأشقاء الثلاثة جرى سحبهما مرتين، أما المنزل الثاني الذي تم

اختياره بالعينة فيسكن فيه رجل وزوجته، أحدهما شقيق الزوجة في البيت الأول والثاني غير مصاب بالمرض $(y_6 = 0)$ ولا يوجد لديه أشقاء في المدينة؛ لذا نرى أن الزوج (الشخص السادس) يشكل شبكة بحجم وحدة واحدة أي $m_3 = 1$. أما المنزل الثالث يسكن فيه شخص واحد بالغ ويحمل المرض $(y_7 = 1)$ ولا يوجد لديه أشقاء في المدينة، وهذا يشكل الشبكة الرابعة بتعددية تساوي 1 أي $m_4 = 1$. جميع المنازل الباقية التي عددها 97 منزل لا يوجد فيها مصابون ولا يوجد لديهم أشقاء مصابون بالمرض. لذا فإن قيم المتغير $y=0$.

الحل:

لحساب المقدّر المتعدد سنقوم بحساب w_j في جميع المنازل. بالنسبة للمنزل الأول فإن $w_1 = \frac{1}{2} + \frac{2}{3} = \frac{7}{6}$. أما بالنسبة للمنزل الثاني فإن $w_2 = \frac{2}{3} + \frac{0}{1} = \frac{2}{3}$.
وأما بالنسبة للمنزل الثالث فإن $w_3 = \frac{1}{1} = 1$. لكل منزل من المنازل الباقية التي عددها 97 فإن $w_j = 0$. الآن نستطيع أن نحسب قيمة المقدّر المتعدد للمجمع الكلي τ

$$\hat{\tau}_m = \frac{5000}{100}\left(\frac{7}{6} + \frac{2}{3} + 1 + 0 + \ldots + 0\right) = 141.7$$

لكي نحسب تقدير التباين إلى $\hat{\tau}_m$ لابد من حساب الوسط الحسابي وتباين العينة للمتغير W، لذا فإن $\bar{w} = 0.02833$ و $s_w^2 = 0.02753$. الآن نستطيع حساب تقدير التباين إلى $\hat{\tau}_m$ فهو

$$s^2(\hat{\tau}_m) = \frac{5000(5000-100)}{100}(0.02753) = 6745.$$

أما تقدير الخطأ المعياري فسيكون

$$\sqrt{s^2(\hat{\tau}_m)} = \sqrt{6745} = 82.13.$$

3.19 المعاينة الشبكية الطبقية

قد تنشأ تعقيدات عندما يتم سحب الوحدات المختارة من مجتمع مؤلف من طبقات وذلك بسبب أن بعض وحدات المشاهدة يمكن أن ترتبط بالوحدات المختارة في أكثر من طبقة؛ لذا فإن المشاهدات في الطبقات المختلفة ليست مستقلة كما هو متعارف عليه في العينة العشوائية الطبقية المعروفة.

لنفترض أن الوحدات المختارة في المجتمع التي عددها M مقسمة إلى L من الطبقات بحجم M_h وحدة لكل طبقة h، ولنفرض أنه جرى سحب عينة عشوائية بسيطة بحجم n_h من كل طبقة حيث إن $h=1,2,\ldots,L$. لكل وحدة اختيار في العينة، جميع وحدات المشاهدة المرتبطة بها يجري ضمها للعينة بغض النظر عن الطبقات التي تنتمي لها هذه الوحدات. لنفترض أن A_{hj} عبارة عن مجموعة من الوحدات المشاهدة التي ترتبط بالوحدة j في الطبقة h. لوحدة المشاهدة i نفرض أن m_i تمثل عدد وحدات الاختيار التي يمكن أن تكون جاءت من أكثر من طبقة ومرتبطة بالوحدة i. لنعرف وحدة الاختيار j في الطبقة h المتغير الجديد الذي نرغب في دراسته

$$w_{hj} = \sum_{i \in A_{hj}} \frac{y_i}{m_i}$$

لنعرف الوسط الحسابي للمتغير w في الطبقة h بما يأتي:

$$\overline{w}_h = \frac{1}{n_h} \sum_{j=1}^{n_h} w_{hj}$$

المقدّر الطبقي المتعدد للمجموع الكلي للمجتمع كما اقترحه Birnbaum and Sirken (1965) هو

$$\hat{t}_m = \sum_{h=1}^{L} M_h \overline{w}_h$$

حيث إنه مقدّر غير متحيز للمجموع الكلي للمجتمع وبتباينه

$$var(\hat{\tau}_m) = \sum_{h=1}^{L} \frac{M_h(M_h - n_h)}{n_h} \sigma_{w_h}^2$$

حيث إن $\sigma_{w_h}^2$ يمثل التباين للمجتمع المحدود لقيم المتغير w في الطبقة h. يمكننا الحصول على تقدير غير متحيز لتباين $\hat{\tau}_m$ أعلاه باستبدال $\sigma_{w_h}^2$ بتباين العينة للمتغير w في الطبقة h وهو s_{wh}^2.

لابد من ملاحظة أنه بالرغم من كون $\hat{\tau}_m$ مقدّراً غير متحيز للمجموع الكلي للمجتمع τ فإن $M_h \overline{w}_h$ قد لا يكون بصورة عامة مقدّراً غير متحيز إلى المجموع الكلي لكل طبقة. وذلك ربما يكون $\overline{w}_h$ متحيزاً بصورة جزئية لقيم المتغير y لوحدات المشاهدة المرتبطة مع طبقات أخرى، على سبيل المثال إذا كانت وحدات الاختيار هي المنازل في منطقة سكنية، والمنطقة مقسمة إلى طبقات حسب الموقع الجغرافي، ووحدات المشاهدة عبارة عن أشخاص بالغين ومرتبطين بعلاقات أخوية مع آخرين في المنطقة السكنية؛ لذا فإن اختيار منزل في طبقة ما قد يؤدي إلى الإبلاغ عن أشقاء موجودين في أكثر من طبقة. نقوم بجمع قيم المتغير y للأشقاء في الطبقات المختلفة في w_{hj} لهذا المنزل. لمعالجة هذه المشكلة ولمزيد من المعلومات يراجع كلٌّ من Birnbaum and Sirken (1965) و Thompson (2002).

References

1. Birnbaum, Z. W. and Sirken, M. G. (1965). Design of Sample Survey to Estimate the Prevalence of Rare Diseases: Three Unbiased Estimators, *Vital and Health Statistics*, Ser. 2, No. 11, Washington, DC: US Government Printing Office.
2. Czaja, R. F. Snowdon, C. B. and Casady, R. J. (1986). Reporting Biased and Sampling Errors in a Survey of a Rare Population using Multiplicity Counting Rules, *Journal of the American Statistical Association*, 81, 411–419.
3. Faulkenberry, G. D. and Garoui, A. (1991). Estimating a Population Total using an Area Frame, *Journal of the American Statistical Association*, 86, 445–449.
4. Kalton, G. and Anderson, D. W. (1986). Sampling Rare Populations, *Journal of the Royal Statistical Society* A, 149, 65–82.
5. Levy, P. S. (1977). Optimum Allocation in Stratified Random Network Sampling for Estimating the Prevalence Attributes in Rare Populations, *journal of the American Statistical Association*, 72, 758–763.
6. Nathan, G. (1976). An Empirical Study of Response and Sampling Errors for Multiplicity Estimates with Different Counting Rules, *Journal of the American Statistical Association*, 71, 808–815.
7. Sirken, M. G. (1970). Household Survey with Multiplicity, *Journal of the American Statistical Association*, 63, 257–266.
8. Sirken, M. G. (1972a). Stratified Sample Survey with Multiplicity, *Journal of the American Statistical Association*, 67, 224–227.
9. Sirken, M. G. (1972b). Variance Components of Multiplicity Estimators, *Biometrics*, 28, 869–873.
10. Sirken, M. G. and Levy, P. S. (1974). Multiplicity Estimation of Proportions Based on Ratios of Random Variables, *Journal of the American Statistical Association*, 69, 68–73.
11. Sudman, S., Sirken, M. G. and Cowan, C. D. (1988). Sampling Rare and Elusive Populations, *Science*, 240, 991–996.
12. Thompson, S. K. (2002). Sampling, 2end Wiley, New York.

الفصل العشرون

معاينة المجموعات المرتبة
Ranked Set Sampling

1.20 مقدمة

تُعَدُّ تكلفة جمع البيانات من أهم أسباب استخدام طرق المعاينة المختلفة، خصوصاً إذا كانت تكاليف قياس صفات معينة للوحدات التي نرغب في دراستها مرتفعةً أو تحتاج وقتاً طويلاً لقياسها. لقد كان McIntyre(1952) أول من اقترح طريقة أكثر فاعلية لتقدير إنتاج حقول الرعي في أستراليا والتي أصبحت تعرف فيما بعد بطريقة معاينة المجموعات المرتبة. تُعَدُّ طريقة معاينة المجموعات المرتبة من الطرق الفاعلة في تقليل كلفة جمع البيانات؛ وذلك من خلال تخفيض حجم العينة إذا توافرت بعض الشروط. وبالرغم من كون معاينة المجموعات المرتبة قديمة إلا أنها لم تستخدم بشكل واسع وتنتشر إلا في العقد الأخير من القرن الماضي على الرغم من فاعليتها في خفض تكاليف جمع البيانات.

إن أول من اقترح فكرة معاينة المجموعات المرتبة هو McIntyre(1952) ضمن جهوده المتميزة لإيجاد مقدّر يكون أكثر فاعلية لتقدير إنتاج حقول الرعي الواسعة في أستراليا؛ لأن قياس إنتاجية الحقول يتطلب قطع ووزن الحشيش الموجود في هذه الحقول وهذا يتطلب جهداً ووقتاً كبيرين، ولكن شخصاً خبيراً يستطيع أن يرتب مجموعة من القطع المحتوية على الحشيش بالعين المجردة حسب كمية إنتاجها من الأدنى إلى الأعلى من غير أن تكون

هناك حاجة إلى قطع ووزن الحشيش في هذه القطع؛ لذا فإن McIntyre تبنى طريقة المعاينة الآتية: نقوم بسحب n من المجموعات بطريقة عشوائية حجم كل مجموعة n من الوحدات (القطع)، ومن ثم نقوم بترتيب كل مجموعة من القطع التي عددها n بالعين المجردة من دون قياس الوحدات من دون تكاليف حسب إنتاجيتها من الحشيش من الأدنى إلى الأعلى. من المجموعة الأولى والمرتبة من الأدنى إلى الأعلى نقوم بقطع ووزن الحشيش، من القطعة التي تحتوي على الحد الأدنى (القطعة بالمرتبة الدنيا) من الحشيش. ومن المجموعة الثانية نقوم بعد ترتيبها من حيث الإنتاج من الأدنى إلى الأعلى بقطع الحشيش من القطعة التي تحتوي على المرتبة الثانية (القطعة بالمرتبة الثانية) من حيث كمية الحشيش. وهكذا إلى أن نقطع الحشيش من القطعة ذات الإنتاج الأعلى ضمن المجموعة الأخيرة (القطعة بالمرتبة العليا). يمكننا أن نكرر العملية r من الدورات أو المرات لنحصل على عينة بحجم nr من الوحدات.

يبدو أن فكرة العينة المرتبة التي اقترحها McIntyre لم تلاقي رواجاً، وبقيت منسية حتى عام 1966 حيث قام Halls and Dell بتطبيق هذه الفكرة لتقدير إنتاج علف الماشية في غابات أشجار الصنوبر. في الحقيقة إن أول من استخدم مصطلح معاينة المجموعات المرتبة هما Halls and Dell. أما أول من قدم البراهين الرياضية لهذا النوع من المعاينة فهما العالمان اليابانيان Takahasi and Wakimoto(1968) حيث برهنا على أن الوسط الحسابي لهذا النوع من المعاينة مقدّر غير متحيز إلى الوسط الحسابي للمجتمع، وتباينه أقل من تباين الوسط الحسابي للعينة العشوائية البسيطة، وذلك إذا كان ترتيب الوحدات يتم بصورة كاملة، أي دون أخطاء في ترتيب وحدات المجموعة من الأدنى إلى الأعلى. أما Dell and Clutter(1972) فقد توصلا إلى نفس النتيجة أعلاه ولكن مع إسقاط الشرط الذي يفترض كون الوحدات داخل المجموعة ترتب بصورة كاملة، أي سمحا للخطأ في ترتيب الوحدات داخل المجموعة أو

عينة من الأصغر إلى الأكبر. وهذا مهم في الحياة العملية حيث يصعب التأكد من عدم وجود أخطاء في ترتيب وحدات المجموعة أو العينة الواحدة. يُعَدُّ Dell and Clutter(1972) وDavid and Levine(1972) أول من أعطى معالجة نظرية لمعاينة المجموعات المرتبة مع الخطأ في ترتيب الوحدات داخل كل مجموعة أو عينة. اقترحت Stokes(1977) باستخدام المتغير الملازم أو المصاحب لتقدير رتب المتغير الذي نرغب في دراسته والذي يصعب ترتيب وحداته بالعين المجردة. قامت Stokes(1980a, b) بتقدير تباين المجتمع ومعامل الارتباط بين متغيرين إذا كان المجتمع يتبع التوزيع الطبيعي الثنائي باستخدام عينة المجموعات المرتبة. قام كلٌّ من Muttlak and McDonald(1990a,b,1992) باستخدام معاينة المجموعات المرتبة إذا تم سحب الوحدات من المجتمع باحتمالات متناسبة مع أحجامها، واقترحا مقدّرات للوسط الحسابي للمجتمع، واثبتا أن تباين هذه المقدّرات أقل من تباين العينة العشوائية البسيطة، كما استخدما الطريقة –أول مرة– مع طريقة المعاينة بخط التقاطع. وأخيراً قاما بتطبيق هذه الطريقة بتقدير التغطية والكثافة لنوع معين من الشجيرات في حقل الدراسة.

شهد العقد الأخير من القرن العشرين تطوراً كبيراً في استخدام معاينة المجموعات المرتبة؛ وذلك من خلال تناول أوجه متعددة لاستخدام هذا النوع من المعاينة. سنشير إلى بعض هذه البحوث بعجالة ولكننا ننصح بمراجعة Patil et al.(1999) وMuttlak and Al-Saleh(2000) للحصول على مزيد من البحوث في هذه الطريقة حتى عام 2000. لتقدير دالة التوزيع الاحتمالي يراجع كلٌّ من Stoke and Sager(1988) وKvam and Samanigeo(1994) وChen(2000a). أما فيما يختص الطرق غير المعلماتية يراجع على سبيل المثال كلٌّ من Bohn and Wolfe(1992,1994) وBohn (1996,1998) وOmer(1999,2002). فيما يخص الطرق المعلماتية باستخدام معاينة المجموعات المرتبة فهي أكثر مما نستطيع ذكره هنا، نذكر منها: Fei et. Al.(1994)

و(Bohj(1997a,b,c,1999a,b و(Abu-Dayyeh and Muttlak(1996
و(Bohj and Ahsanullah(1996 و(Shen(1994 و(Stokes(1995 و(Chen(2000b
و(1996).Sinha et al لمزيد من المعلومات ومراجعة حول الطرق المعلماتية يراجع (Ni and Sinha(1998. لقد لقي تحليل الانحدار باستخدام طريقة معاينة المجموعات المرتبة لجمع البيانات اهتمام مجموعة من الباحثين من بينهم (Muttlak(1995,1996b,1998 و(1993).Patil et al و(Yu and Lam(1997 و(Chen(2001. أما التصميم الأمثل في سياق معاينة المجموعات المرتبة في حالة كون المجموعات غير متساوية فقد تناولها (1997).Kaur et al و(Chen(2001.
أما استخدام معاينة المجموعات المرتبة في مراقبة وضبط الجودة فقد تناولها كلٌّ من الباحثين (Sinha(1997 و Salazar and و Muttlak and Al-Sabah(2003a,b).

2.20 تقدير الوسط الحسابي للمجتمع

يمكن تلخيص طريقة معاينة المجموعات المرتبة كما هو موضح في الشكل 1 أدناه، ففي الخطوة الأولى نقوم بسحب 5 مجموعات (عينات) بطريقة عشوائية من المجتمع، حجم كل مجموعة 5 وحدات. في الخطوة الثانية نقوم بترتيب كل مجموعة من المجموعات الخمسة من الأدنى إلى الأعلى، وفي الخطوة الثالثة نقوم بقياس الوحدات الواقعة على القطر، وبذلك نكون حصلنا على العينة العشوائية المرتبة وبحجم 5 وحدات. بإمكاننا إعادة العملية من جديد للحصول على 5 وحدات جديدة وهكذا.

1.2.20 تقدير الوسط الحسابي للمجتمع إذا كان ترتيب الوحدات داخل لكل مجموعة دون أخطاء

لنفترض أن $X_{(i:n)j}$ القيمة الإحصائية للمتغير برتبة i في العينة التي حجمها n في الدورة j حيث إن $i=1,2,\ldots,n$ و $j=1,2,\ldots,r$. لذا فإن المقدّر غير المتحيز لتقدير

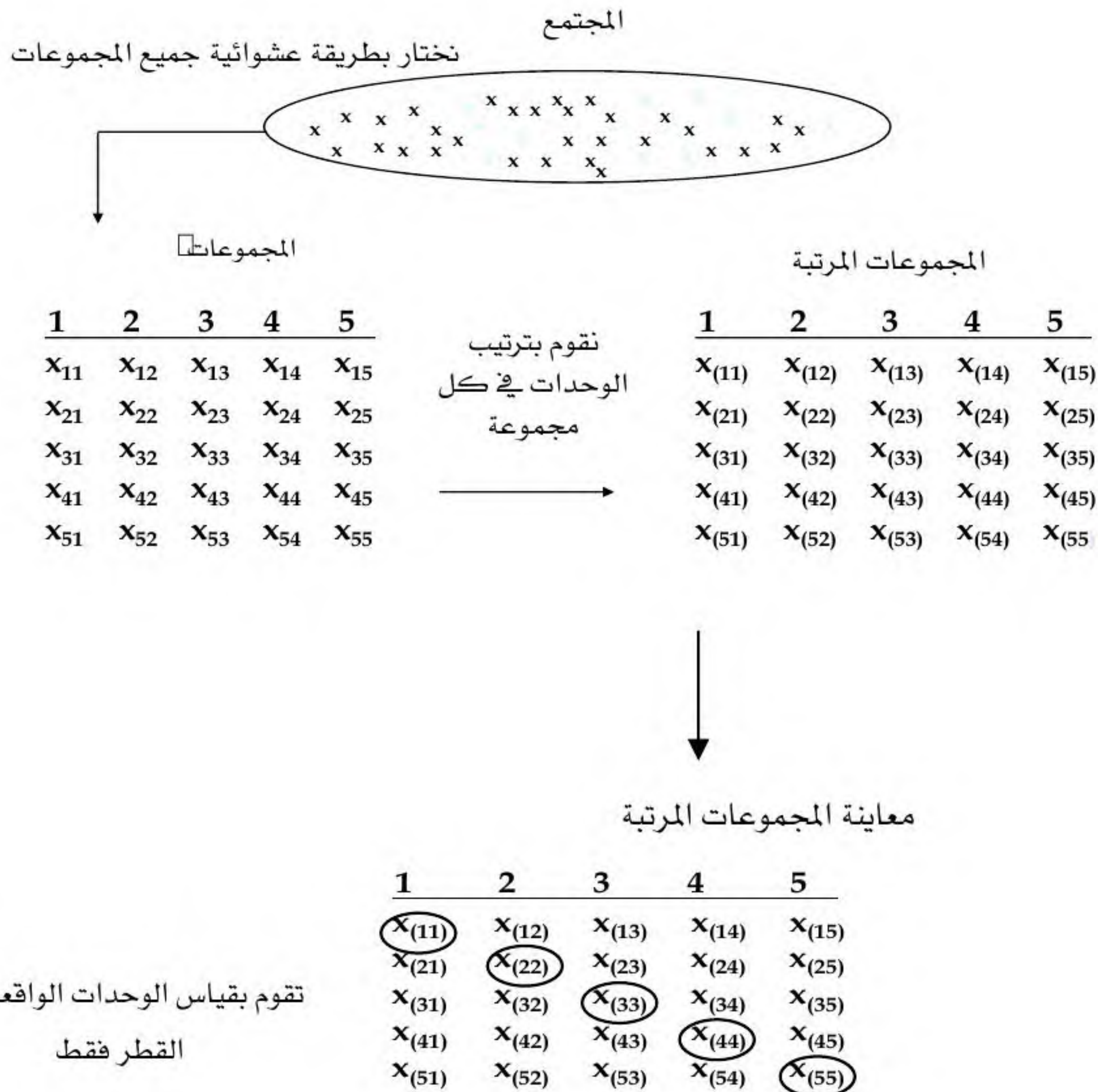
متوسط المجتمع إذا كان ترتيب الوحدات داخل كل مجموعة كاملاً أي دون أخطاء هو

$$\bar{X}_{rss} = \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} X_{(i:n)\,j}$$

أما تباين $\bar{X}_{rss}$ فهو

$$var\left(\bar{X}_{rss}\right) = \frac{1}{n^2 r}\sum_{i=1}^{n}\sigma^2_{(i:n)}$$

الشكل 1 : خطوات اختيار عينة المجموعات المرتبة



حيث إن

$$\sigma^2_{(i:n)} = E\left[X_{(i:n)} - E\left(X_{(i:n)}\right)\right]^2$$

لقد برهنا Takahasi and Wakimoto (1968) على أن $\bar{X}_{rss}$ هو مقدّر غير متحيز لتقدير متوسط المجتمع وعلى أن

$$var(\bar{X}_{rss}) \le var(\bar{X}_{srs})$$

حيث إن $\bar{X}_{srs}$ يمثل متوسط العينة العشوائية البسيطة وبالحجم نفسه، أما تقدير التباين فيمكن الحصول على مقدّر غير متحيز إلى $var(\bar{X}_{rss})$ وهو

$$s^2_{\bar{X}_{rss}} = \frac{1}{n^2 r(r-1)} \sum_{j=1}^{r} \sum_{i=1}^{n} \left(X_{(i:n)j} - \bar{X}_{(i:n)}\right)^2$$

حيث إن

$$\bar{X}_{(i:n)} = \frac{1}{r} \sum_{j=1}^{r} X_{(i:n)j} \quad i = 1, 2, \ldots, n$$

يراجع كلٌّ من Johnson et al. (1993) Stokes(1980) وللبراهين الرياضية ومزيد من المعلومات.

أما إذا كانت عدد الدورات $r=1$ فإن $\bar{X}_{rss}$ سيعرف كما يأتي:

$$\bar{X}_{rss} = \frac{1}{n} \sum_{i=1}^{n} X_{(i:n)}$$

و أما تباينه فسيكون

$$var\left(\bar{X}_{rss}\right) = \frac{1}{n^2} \sum_{i=1}^{n} \sigma^2_{(i:n)}$$

2.2.20 تقدير الوسط الحسابي إذا كان ترتيب الوحدات ضمن كل مجموعة فيها أخطاء

لنفترض أن $X_{[i:n]j}$ القيمة الإحصائية للمتغير برتبة المقدرة i في العينة التي حجمها n في الدورة j حيث إن $i=1, 2, \ldots, n$ و $j=1,2,\ldots,r$. ونعني بالرتبة المقدرة هنا أن هذه الرتبة يمكن أن تكون خطأ؛ لذا فإن المقدّر غير المتحيز لتقدير لمتوسط المجتمع إذا كان ترتيب الوحدات داخل كل مجموعة فيها أخطاء هو

$$\bar{X}_{rsse} = \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} X_{[i:n]j}$$

أما تباين $\bar{X}_{rsse}$ فهو

$$var\left(\bar{X}_{rsse}\right) = \frac{1}{n^2 r}\sum_{i=1}^{n}\sigma^2_{[i:n]}$$

حيث إن

$$\sigma^2_{[i:n]} = E\left[X_{[i:n]} - E\left(X_{[i:n]}\right)\right]^2$$

لقد برهنا Dell and Clutter(1972) على أن $\bar{X}_{rss}$ هو مقدّر غير متحيز لتقدير متوسط المجتمع، وعلى أن

$$var(\bar{X}_{rsse}) \leq var(\bar{X}_{srs})$$

بغض النظر فيما إذا كان هناك أخطاء بترتيب وحدات المجموعة أم لا.

أما تقدير التباين فيمكن الحصول على مقدّر غير متحيز إلى $var(\bar{X}_{rsse})$ وهو

$$s^2_{\bar{X}_{rsse}} = \frac{1}{n^2 r(r-1)}\sum_{j=1}^{r}\sum_{i=1}^{n}(X_{[i:n]j} - \bar{X}_{[i:n]})^2$$

حيث إن

$$\bar{X}_{[i:n]} = \frac{1}{r}\sum_{j=1}^{r} X_{[i:n]j} \quad i = 1, 2, \ldots, n$$

مثال: لتقدير الجزء غير المملوء من القناني الزجاجية لمشروب البيبسي كولا، قمنا باستخدام معاينة المجموعات المرتبة، حيث تم سحب 16 زجاجة بصورة عشوائية من خط إنتاج المصنع الكائن في مدينة الخبر في المملكة العربية السعودية، ومن ثم قمنا بتقسيم هذه القناني الست عشرة إلى أربع مجموعات، كل مجموعة تحتوي على $n=4$ قنينات، وبالعين المجردة حُدِدَت القنينة التي تحتوي على أقل ارتفاع من الجزء الفارغ وقياسه من المجموعة الأولى. من المجموعة الثانية حُددت، القنينة التي تحتوي على ثاني أقل ارتفاع فارغ وقياسه. وهكذا حتى وصلنا إلى المجموعة الرابعة حيث قمنا بقياس أعلى ارتفاع فارغ من بين قناني المجموعة الرابعة، ثم قمنا بإعادة الدورة $r=8$ مرات للحصول على عينة بحجم $nr=32$ قنينة. قَدّر متوسط ارتفاع الجزء الفارغ لإنتاج المصنع من قناني البيبسي كولا وخطئه المعياري.

المجموعات

4	3	2	1	الدورات
6.0	6.1	5.8	5.7	1
6.1	5.8	5.8	5.6	2
5.9	6.1	5.8	5.6	3
6.7	6.1	6.0	5.8	4
6.0	6.0	6.0	5.8	5
6.0	6.1	5.9	5.8	6
6.1	5.9	5.8	5.9	7
6.0	6.1	5.9	5.9	8

الحل: متوسط ارتفاع الجزء الفارغ هو

$$\bar{X}_{rss} = \frac{1}{4(8)}\sum_{i=1}^{4}\sum_{j=1}^{8} X_{(i:n)\,j} = 5.94$$

لحساب تقدير تباين $\bar{X}_{rss}$ ، نحتاج أن نحسب الوسط الحسابي $\bar{X}_{(i:n)}$ لكل مجموعة، وكذلك مجموع المربعات الفروق داخل كل مجموعة، الجدول الآتي يعطينا هذه المعلومات لكل مجموعة

المجموعة	1	2	3	4
$\bar{X}_{(i:n)}$	5.763	5.875	6.025	6.10
$\sum_{j=1}^{8}(X_{(i:n)\,j}-\bar{X}_{(i:n)})^2$	0.09875	0.05500	0.09499	0.43999

$$s^2_{\bar{X}rss} = \frac{1}{4^2 8(8-1)}\sum_{j=1}^{8}\sum_{i=1}^{4}(X_{(i:n)\,j}-\bar{X}_{(i:n)})^2 = \frac{0.688748}{896} = 0.00076869$$

لذا فإن الخطأ المعياري إلى $\bar{X}_{rss}$ هو

$$s_{\bar{X}rss} = \sqrt{s^2_{\bar{X}rss}} = \sqrt{0.00076869} = 0.0277$$

3.20 طرق معدلة لمعاينة المجموعات المرتبة

إن ترتيب الوحدات داخل المجموعات من الأصغر إلى الأكبر بالنسبة للمتغير الذي نرغب في دراسته بالعين المجردة غالباً ما يكون تنفيذه صعباً إذا ما كان حجم المجموعة كبيراً (أكثر من خمس وحدات)، وإذا أنجز بالغالب ستعتري عملية الترتيب أخطاء، وهذا سيؤدي إلى خفض فاعلية معاينة المجموعات المرتبة؛ لذا أصبح من الضرورة البحث عن بدائل لعملية ترتيب

الوحدات داخل المجموعة لتلافي الأخطاء في الترتيب. سنتناول هنا ثلاث طرق بديلة من بين الطرق المقترحة لتعديل عملية ترتيب الوحدات داخل المجموعات.

1.3.20 معاينة المجموعات المرتبة وسطياً

تُعَدُّ معاينة المجموعات المرتبة وسطياً من الطرق الفاعلة في تقليل الأخطاء في عملية ترتيب الوحدات داخل المجموعات، ويمكن تلخيص الطريقة فيما يأتي: نقوم بسحب n من المجموعات (العينات) بطريقة عشوائية من المجتمع حجم كل مجموعة n من الوحدات. ثم نقوم بترتيب المجموعات من الأصغر إلى الأكبر، إذا كان حجم المجموعة n عبارة عن عدد فردي، فإننا نختار الوسيط من كل مجموعة للقياس، أي الوحدة ذي الرتبة $(n+1)/2$، أما إذا كان حجم المجموعة أو المجموعات عدداً زوجياً، فإننا نقوم بسحب الوحدات ذات الرتبة $n/2$ للقياس من نصف المجموعات، ومن النصف الثاني نقوم بسحب الوحدات ذات الرتبة $(n/2)+1$ وقياسها، في كلا الحالتين سنحصل على عينة بحجم n من الوحدات. يمكننا إعادة الدورة r من المرات للحصول على عينة بحجم nr.

لنفترض أن $X_{(i:m)j}$ القيمة الإحصائية للمتغير للوسيط في العينة i والتي حجمها n في الدورة j حيث إن $i=1,2,\ldots,n$ و $j=1,2,\ldots,r$ إذا كان حجم المجموعة أو العينة عدداً فردياً. أما إذا كان حجم المجموعة عبارة عن عدد زوجي فإن i تمثل المرتبة الإحصائية $n/2$ لقيمة المتغير في المجموعة والتي بحجم n من الوحدات حيث إن $i=1,2,\ldots L=n/2$. ويمثل المرتبة الإحصائية $(n/2)+1$ لقيمة المتغير في لمجموعة حيث إن $i=L+1,L+2,\ldots,n$. المقدّر لتقدير متوسط المجتمع باستخدام معاينة المجموعات المرتبة وسطياً هو

$$\bar{X}_{mrss} = \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} X_{(i:m)j}$$

أما تباين $\bar{X}_{mrss}$ فهو

$$\operatorname{var}\left(\bar{X}_{mrss}\right)=\frac{1}{n^{2}r}\sum_{i=1}^{n}\sigma_{(i:m)}^{2}$$

حيث إن

$$.\ \sigma_{(i:m)}^{2}=E\left[X_{(i:m)}-E\left(X_{(i:m)}\right)\right]^{2}$$

إذا كان المجتمع متماثلاً حول الوسط الحسابي $\bar{X}_{mrss}$ هو مقدّر غير متحيز لتقدير متوسط المجتمع، وكذلك فإن

$$\operatorname{var}(\bar{X}_{mrss})\leq \operatorname{var}(\bar{X}_{rss})\leq \operatorname{var}(\bar{X}_{srs})$$

لمزيد من المعلومات يراجع Muttlak(1997).

2.3.20 معاينة المجموعات المرتبة تطرفياً

لقد اقترح أكثر من باحث استخدام معاينة المجموعات المرتبة تطرفياً لتسهيل عملية ترتيب الوحدات داخل المجموعات، ولتقليل أخطاء ترتيب الوحدات الذي يؤدي إلى تقليل فاعلية معاينة المجموعات المرتبة. يمكننا تلخيص الطريقة فيما يأتي: نقوم بسحب n من المجموعات (العينات) بطريقة عشوائية من المجتمع حجم كل مجموعة n من الوحدات. ثم نقوم بترتيب المجموعات من الأصغر إلى الأكبر، إذا كان حجم المجموعة أو المجموعات عدداً زوجياً، فإننا نقوم بسحب الوحدة ذي الرتبة الأصغر من $n/2$ مجموعة للقياس أي نصف المجموعات، ومن النصف الثاني تختار الوحدات ذات الرتبة العليا ونقيسها، أما إذا كان حجم المجموعة n عبارة عن عدد فردي، فإننا نختار $(n-1)/2$ من المجموعات ونقوم بقياس الوحدات ذات الرتبة الصغرى، ومن $(n-1)/2$ نختار الوحدات ذات الرتبة العليا ونقيسها. وأخيراً نختار إحدى المجموعات ونقيس الوسيط من هذه المجموعة بعد ترتيب وحداتها. في كلتا

الحالتين سنحصل على عينة بحجم n من الوحدات، ويمكننا إعادة الدورة r من المرات للحصول على عينة بحجم nr.

لنفترض أن $X_{(i\square e)j}$ القيمة الإحصائية للمتغير بالرتبة الصغرى i من العينة والتي حجمها n في الدورة j حيث $L_1 = n/2$,..$i=1,2,\ldots$ و $j=1,2,\ldots r$ إذا كان حجم المجموعة أو العينة عدداً زوجياً، كذلك تمثل القيمة الإحصائية للمتغير بالرتبة العليا من العينة التي حجمها n في الدورة j حيث $i = L_1+1, L_1+1, \ldots, n$. أما إذا كان حجم المجموعة عبارة عن عدد فردي فإن i تمثل المرتبة الإحصائية الصغرى لقيمة المتغير من المجموعة بحجم n و $i=1,2,\ldots, L_2=(n-1)/2$ والوسيط من إحدى المجموعات والمرتبة العليا لقيم المتغير من المجموعة بحجم n و $i=L_2+2, L_2+3,\ldots, n$. يمكننا إعادة الدورة r من الدورات للحصول على عينة بحجمnr. المقدّر لتقدير متوسط المجتمع باستخدام معاينة المجموعات المتطرفة هو

$$\bar{X}_{erss} = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{r} X_{(i\square e)j}$$

أما تباين $\bar{X}_{erss}$ فهو

$$var(\bar{X}_{erss}) = \frac{1}{n^2 r}\sum_{i=1}^{n}\sigma^2_{(i\square e)}$$

حيث إن

$$\sigma^2_{(i\square e)} = E\ [X_{(i\square e)} - E(X_{(i\square e)})]^2$$

هناك أكثر من طريقة لاختيار الوحدات باستخدام معاينة المجموعات المرتبة تطرفياً، كذلك هنالك أكثر من صيغة لمقدّر متوسط المجتمع. يراجع Stkose(1980) و Samawi et al.(1996) و Muttlak and Al-Sabah(2003a) لمزيد من المعلومات.

3.3.20 معاينة المجموعات المرتبة باستخدام المتغير المصاحب

لنفترض أن المتغير الذي نرغب في دراسته X يصعب قياسه وترتيبه أي ترتيب مجموعة من الوحدات حسب قيمة المتغير X ولكن هناك متغير Y مصاحب ومرتبط مع المتغير X يسهل ترتيبه. يمكننا أن نستخدم المتغير Y لتقدير رتب المتغير X على النحو الآتي: نقوم بسحب n من المجموعات حجم كل مجموعة n من الوحدات الثنائية أي تحتوي المتغيرين X وY. نرتب المجموعة الأولى باستخدام المتغير Y ونقوم بقياس المتغير X المصاحب للرتبة الصغرى للمتغير Y. من المجموعة الثانية نقوم بقياس المتغير X المصاحب للرتبة الثانية للمتغير Y. وهكذا إلى أن نقيس المتغير X المصاحب للرتبة العليا للمتغير Y من المجموعة الأخيرة. يمكننا إعادة الدورة r من المرات لنحصل على عينة بحجم nr من الوحدات. لابد من ملاحظة أن رتب المتغير X ستكون مع أخطاء بالترتيب أي $X_{[i:n]j}$ حيث تمثل القيمة الإحصائية للمتغير برتبة المقدرة i في العينة والتي حجمها n في الدورة j، حيث إن $i=1,2,\dots n$ و $j=1,2,\dots r$.

لنفترض أن (Y,X) تتبع التوزيع الثنائي الطبيعي وعلاقة الانحدار بين X وY خطية. لذا يمكننا أن نتبع طريقة Stokes(1977) ونكتب

$$X=\mu_x+\ \rho\frac{\sigma_x}{\sigma_y}(Y-\mu_y)+\varepsilon$$

حيث إن Y و ε مستقلين والمتوسط والتباين للمتغير ε هما على النحو الآتي 0 و $\sigma_x^2(1-\rho^2)$ ، أما ρ فيمثل معامل الارتباط بين X وY و $\mu_x,\mu_y,\sigma_x,\sigma_y$ يمثلون متوسط والانحراف المعياري للمجتمع للمتغيرين X وY. لذا فإن المقدّر غير المتحيز لمتوسط المجتمع للمتغير X هو

$$\overline{X}_{rssc}=\frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r}X_{[i:n]j}$$

أما التباين للمقدّر $\overline{X}_{rssc}$ فيمكن الحصول عليه من Stokes(1977) وهو

$$var(\overline{X}_{rssc}) = \frac{\sigma_x^2}{nr}[(1-\rho^2)+\frac{\rho^2}{n\sigma_y^2}\sum_{i=1}^{n}\sigma_{y(i:n)}^2]$$

حيث إن

$$\sigma_{y(i:n)}^2 = E\ [Y_{(i:n)} - E(Y_{(i:n)})]^2$$

4.20 تقدير الوسط الحسابي باستخدام الانحدار لمعاينة المجموعات المرتبة

لقد حاول أكثر من باحث الجمع بين معاينة المجموعات المرتبة وخط الانحدار نذكر منهم Barreto and Barnett(1999) و Barnett and Moore (1977) و Muttlak(1995,1998) و Patil et al.(1993) و Yu and Lam(1997).

سوف نقتصر في بحثنا هنا على بعض النتائج التي اقترحها Yu and Lam (1997) والتي لها علاقة مباشرة بتقدير الوسط الحسابي للمتغير الذي نرغب في دراسته.

سوف نتبع نفس الطريقة التي اقترحتها Stokes (1977) والتي تناولناها أعلاه. ونكتب

$$X = \mu_x + \rho\frac{\sigma_x}{\sigma_y}(Y-\mu_y) + \varepsilon$$

لنفترض أن $Y_{(i:n)j}$ و $X_{[i:n]j}$ تمثل القيمة الإحصائية للمتغير برتبة i في العينة والتي حجمها n في الدورة j للمتغير Y والقيمة الإحصائية للمتغير برتبة المقدرة i في العينة والتي حجمها n في الدورة j للمتغير X، نستطيع أن نكتب

$$X_{[i:n],j} = \mu_x + \frac{\rho\sigma_x}{\sigma_y}(Y_{(i:n),j} - \mu_y) + \varepsilon_{ij};\quad i=1,2,\dots,n;\ j=1,2,\dots,r$$

لقد اقترح Yu and Lam (1997) مقدّراً غير متحيز لتقدير متوسط للمتغير X بافتراض أن العلاقة

بين X وY خطية وهو

$$\bar{X}_{reg} = \bar{X}_{rsse} + \hat{\beta}(\mu_y - \bar{Y}_{rss})$$

حيث إن

$$\hat{\beta} = \frac{\sum_{i=1}^{n}\sum_{j=1}^{r}(Y_{(i:n)j} - \bar{Y}_{rss})(X_{[i:n]j} - \bar{X}_{rsse})}{\sum_{i=1}^{n}\sum_{j=1}^{r}(Y_{(i:n)j} - \bar{Y}_{rss})^2}$$

أما تباين $\bar{X}_{reg}$ فهو

$$var(\bar{X}_{reg}) = \frac{\sigma_y^2}{nr}(1-\rho^2)[1+E(\frac{\bar{Z}_{rss}}{S_z^2})]$$

حيث إن

$$Z_{(i:n)j} = \frac{Y_{(i:n)} - \mu_y}{\sigma_y}$$

و

$$\bar{Z}_{rss} = \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r} Z_{(i:n)j}$$

و

$$S_z^2 = \frac{1}{nr}\sum_{i=1}^{n}\sum_{j=1}^{r}(Z_{(i:n)j} - \bar{Z}_{rss})^2 .$$

نلاحظ أن $\bar{X}_{reg}$ دائماً يكون مقدّراً غير متحيز إلى الوسط الحسابي للمجتمع بغض النظر عن التوزيع الذي يتبعه المتغيرX ولكن التباين يعتمد على التوزيع الاحتمالي للمتغيرX. لمزيد من المعلومات ولمقارنة المقدّر $\bar{X}_{reg}$ ببعض المقدّرات المعروفة يراجع Yu and Lam(1997).

5.20 تقدير التباين باستخدام معاينة المجموعات المرتبة

لم تكن معاينة المجموعات المرتبة فاعلة بتقدير تباين المجتمع σ^2 كما هو الحال إلى تقدير الوسط الحسابي للمجتمع، لقد أثبتت الدراسة التي قامت بها Stokes (1980 b) أن استخدام معاينة المجموعات المرتبة غير فاعلٍ بتقدير تباين المجتمع بالمقارنة إلى المعاينة التقليدية وهي المعاينة العشوائية البسيطة. وكذلك المقدّر الذي يستخدم معاينة المجموعات المرتبة فإنه مقدّر متحيز، على خلاف المقدّر الذي يستخدم العينة العشوائية البسيطة وهو تباين العينة.

لقد أثبتت Stokes (1980 b) أن المقدّر الذي اقترحته والذي يستخدم معاينة المجموعات المرتبة يمكن أن يكون مقدّراً غير متحيز إذا كان حجم العينة كبيراً وبالتحديد إذا كان بالإمكان إعادة الدورة مرات كثيرة. كذلك يكون هذا المقدّر أكثر فاعلية من تباين العينة العشوائية البسيطة المقدّرة ولكن هذا ليس كبيرًا.

لقد اقترح كلٌّ من Abu‑Dayyeh and Muttlak (2002) مجموعة من المقدّرات لتقدير تباين المجتمع، باستخدام البيانات من معاينة المجموعات المرتبة إذا كان التوزيع الاحتمالي للمجتمع الذي نرغب في دراسته يكون من عائلة التوزيعات الاحتمالية $f(x, \theta, \lambda) = g((x-\theta)/\lambda)/\lambda$. ولقد برهنوا أن معظم هذه المقدّرات هي غير متحيزة لتقدير تباين المجتمع وأكثر فاعلية من تباين العينة العشوائية البسيطة لمعظم التوزيعات التي دُرست من قبل الباحثين.

References

1. Abu-Dayyeh, W. and Muttlak, H. A. (1996). Using Ranked Set Sampling for Hypothesis Tests on the Scale Parameter for the Exponential and Uniform Distributions, *Pakistan Journal of Statistics*, 12, 131-138.
2. Al-Saleh, M. F. and Muttlak, H. A. (1998). A Note on Bayesian Estimation using Ranked Set Sample, *Pakistan Journal of Statistics*, 14, 49-56.
3. Al-Saleh, M. F. and Al-Kadiri, M. A. (1999). Double Ranked Set Sampling, *Statistics and Probability Letters*, 48, 205-212.
4. Al-Saleh, F. M., Al-Shrafat, K. and Muttlak, H. A. (2000). Bayesian Estimation using Ranked Set Sampling, *Biometrical Journal*, 42, 3, 1-12.
5. Barabesi, L. (1998). The Computation of the Distribution of the Sign Test Statistic for Ranked-Set Sampling, *Commun. Statist. Simula.* 27, 833-842.
6. Barnett, V. and Moore, K. (1997). Best Linear Unbiased Estimates in Ranked-Set Sampling with Particular Reference to Imperfect Ordering, *Journal of Applied Statistics*, 24, 697-710.
7. Barnett, V. (1999). Ranked Set Design for Environmental Investigations, *Environmental and Ecological Statistics*, 6, 59-74.
8. Barreto, M. C. M. and Barnett, V. (1999). Best Linear Unbiased Estimators for the Simple Linear Regression Model using Ranked Set Sampling, *Environmental and Ecological Statistics*, 6, 119-133.
9. Bhoj, D. S. and Ahsanullah, M. (1996). Estimation of Parameters of the Generalized Geometric Distribution using Ranked Set Sampling, *Biometrics*, 52, 685-694.
10. Bhoj, D. S. (1997a). Estimation of Parameters of the Extreme Value Distribution using Ranked Set Sampling, *Commun. Statist. Theory Meth.*, 26, 653-667.
11. Bhoj, D. S. (1997b). New Parametric Ranked Set Sampling, *Applied Statistical Sciences*, 6), 275-289.
12. Bhoj, D. S. (1997c). Estimation of Parameters using Modified Ranked Set Sampling, *Applied Statistical Sciences*, 6, 145-163.
13. Bhoj, D. S. (1999a). Estimation of Parameters of the Exponential Distribution using Three Ranked Set Sampling Procedures, *Applied Statistical Sciences*, 8, 155-172.
14. Bhoj, D. S. (1999b). Minimum Variance Linear Unbiased Estimators of the Rayleigh Parameter based on Ranked Set Sampling Procedures, *Applied Statistical Sciences*, 8, 269-277.
15. Bohn, L. L. and Wolfe, D. A. (1992). Nonparametric Two-Sample Procedures for Ranked Set Samples Data, *Journal of American Statistical Association*, 87, 552-561.
16. Bohn, L. L. and Wolfe, D. A. (1992). The Test of Imperfect Rankings on Properties of Procedures based on the Ranked-Set Samples Along with Mann-Whiteny-Wlicoxon Statistics, *Journal of American Statistical Association*, 89, 168-176.
17. Bohn, L. L. (1996). A Review of Nonparametric Ranked-Set Sampling Methodology, *Commun. Statist. Theory Meth.* 25, 2675-2685.
18. Bohn, L. L. (1998). A Ranked-Set Sample Signed-Ranked Statistic, *J Nonparametric Statistics*, 9, 295-306.
19. Chen, Z (1999). Density Estimation using Ranked-Set Sampling Data, *Environmental and Ecological Statistics*, 6, 135-146.
20. Chen, Z (2000a). On Ranked Set Sampling Quantiles and their Applications, *Journal of Statistical Planning and Inference*, 83, 125-135.
21. Chen, Z (2000b). The Efficiency of Ranked-Set Sampling Relative to Simple Random Sampling under Multi-Parameter Families, *Statistica Sinica*, 10, 125-135.
22. Chen, Z (2001). Ranked-Set Sampling with Regression Type Estimators, *Journal of Statistical Planning and Inference*, 92, 181-192.
23. David, H. A. and Levine, D. N. (1972). Ranked Set Sampling in the Present of the Judgment Error, *Biometrics*, 28, 553-55
24. Dell, D. R. and Clutter, J. L. (1972). Ranked Set Sampling Theory with Order Statistics Background, *Biometrics*, 28, 545-53.

25. Fei, H., Sinha, B. K. and Wu, Z. (1994). Estimation of Parameters in Two-Parameter Weibull and Extreme-Value Distributions using Ranked Set Sampling, *Journal of Statistical Research*, 28, 149-161.
26. Halls, L.S. and Dell, T. R. (1966). Trial of Rank Set Sampling for Forage Yields, *Forest Science*, 12, 22-26.
27. Hettmansperger, T. (1995). The Ranked-set Sample Sign Test, *Nonparametric Statistics*, 4, 263-270.
28. Hossain, A. S. and Muttlak, H. A. (1999). Paired Ranked Set Sampling: A More Efficient Procedure, *Enviornmetrics*, 10, 195-212.
29. Johnson, G. D., Patil, G. P. and Sinha, A. K. (1993). Ranked Set Sampling for Vegetation Research, *Abstracta Botanica*, 17, 87-102.
30. Kaur, A., Patil, G. P., Sinha, B. K. and Taillie, C. (1995). Ranked Set Sampling: an Annotated Bibliography, *Environmental and Ecological Statistics*, 2, 25-54.
31. Kaur, A, Patil, G. P. and Taillie, C. (1997). Unequal Allocation Models for Ranked Set Sampling with Skew Distributions, *Biometrics*, 53, 123-130.
32. Kim, Y. H and Arnold, B. C. (1999). Parameter Estimation under Generalized Ranked Set Sampling, *Statistics & Probability Letters*, 42, 353-360.
33. Kvam, P. H. and Samaniego, F. J. (1994). Nonparametric Maximum Likelihood Estimation Based on Ranked Set Samples, *Journal of American Statistical Association*, 89, 526-537.
34. Lam, K., Sinha, B. K. and Wu, Z. (1994). Estimation of Parameters in Two-Parameter Exponential Distribution using Ranked Set Sample, *Annals of the Institute of Statistical Mathematics*, 46, 723-736.
35. Lam, K., Sinha, B. K. and Wu, Z. (1995). Estimation of Location and Scale parameters of a Logistic Distribution using Ranked Set Sample. In: Nagaraja, Sen and Morrison, ed., *Papers in Honor of Herbert A. David*, 187-197.
36. Li, D. Sinha, B. K. and Perron, F. (1999). Random selection in Ranked Set Sampling and its Applications, *Journal of Statistical Planning and Inference*, 76, 185-201.
37. McIntyre, G. A. (1952). A Method of Unbiased Selective Sampling, Using Ranked sets, *Australian Journal of Agricultural Research*, 3, 385-390.
38. Muttlak, H. A. and McDonald, L. L. (1990a). Ranked Set Sampling with Respect to Concomitant Variables and with Size Biased Probability of Selection, *Communication in Statistics Theory and Methods*, 19, 205-219.
39. Muttlak, H. A. and McDonald, L. L. (1990b). Ranked Set Sampling with Size Biased Probability of Selection, *Biometrics*, 46, 435-445.
40. Muttlak, H. A. and McDonald, L. L. (1992). Ranked Set Sampling and Line Intercept Method: A More Efficient Procedure, *The Biometrical Journal*, 34, 329-346.
41. Muttlak, H. A. (1995). Parameters Estimation in a Simple Linear Regression using Ranked Set Sampling, *Biometrical Journal*, 37, 799-810.
42. Muttlak, H. A. (1996a). Estimation of Parameters for One-Way Layout with Ranked Set Sampling, *Biometrical Journal*, 38, 507-515.
43. Muttlak, H. A. (1996b). Estimation of Parameters in a Multiple Regression Model Using Rank Set Sampling, *Information & Optimization Sciences*, 17, 521-533.
44. Muttlak, H. A. (1996c). Pair Ranked Set Sampling, *Biometrical Journal*, 38, 897-885.
45. Muttlak, H. A. (1997). Median Ranked Set Sampling, *Journal of Applied Statistical Science*, 6, 245-55.
46. Muttlak, H. A. (1998). Median Ranked Set Sampling with Concomitant Variables and Comparison with Ranked Set Sampling and Regression Estimators, *Enviornmetrics*, 9, 255-267.
47. Muttlak, H. A and Abu-Dayyeh, W. (1998). Testing Some Hypotheses about the Normal Distribution Using Ranked Set Sampling: A More Powerful Test, *Journal of Information & Optimization Sciences*, 19, 1-11.
48. Muttlak, H. A. (1999a). Median Ranked Set Sampling with Size Biased Probability of Selection, *Biometrical Journal*, 40, 455-465.
49. Muttlak, H. A. (1999b). On Extreme Ranked Set Sampling with Size Biased Probability of Selection. *Far East Journal Theory of Statistics*, 3, 319-329.

50. Muttlak, H. A. and Al□Saleh, M. F. (2000). Recent Developments on Ranked Set Sampling, *Pakistan Journal of Statistics*, 16, 269□290.
51. Abu□Dayyeh, W. A. and Muttlak, H. A. (2002). Variance Estimation for the Locatio Scale Family Distributions using Ranked Set Sampling
52. , *Pakistan Journal of Statistics*, 18.
53. Muttlak, H. A. and Al□Sabah, W. (2003a). Statistical Quality Control using Ranked Set Sampling, *Journal of Applied Statistics*, 30, 1055□1078.
54. Muttlak, H. A. Al□Sabah, W. (2003b). Statistical Quality Control based on Pair and Selected Ranked Set Sampling, *Pakistan Journal of Statistics*, 19, 107□128.
55. Ni Chuiv, N. N. and Sinha, B. K. (1998). On Some Aspects of Ranked Set Sampling in Parametric Estimation, Handbook of statistics, 17, 337□377.
56. Özturk, Ö. (1999). One□and Two□Sample Sign Tests for Ranked Set Sample Selective Designs, *Commun. Statist. Theory Meth.* 28, 1231□1245.
57. Özturk, Ö. (2002). Rank Regression in Ranked□Set Samples, *Journal of American Statistical Association*, 97, 1180□1191.
58. Özturk, Ö. and Wolfe, D. A (2000). Alternative Ranked Set Sampling Protocols for Sign Test, *Statistics & Probability Letters*, 47, 15□23.
59. Patil, G. P., Sinha, A. K. and Taillie, C. (1993). Relative Precision of Ranked set Sampling□ A Comparison with the Regression Estimator, *Enviornmetrics*, 4, 399□412.
60. Patil, G. P., Sinha, A. K. and Taillie, C. (1994). Ranked Set Sampling, *A Handbook of Statistics*, 12,167□200.
61. Patil, G. P., Sinha, A. K. and Taillie, C. (1995). Finite Population Corrections for Ranked Set Sampling, *Annals of the Institute of Statistical Mathematics*, 47, 621□636.
62. Patil, G. P., Sinha, A. K. and Taillie, C. (1999). Ranked Set Sampling□ a Bibliography, *Environmental and Ecological Statistics*, 6, 91□98.
63. Salazar, R. D. and Sinha, A. K. (1997). Control chart X bar based on ranked set sampling. Comunicacion Tecica No. 1□97□09(PE/CIMAT), Department of Probability and Statistics, Centro de Investigation en Matematicas (CIMAT), Apdo. Postal. Gto. 36000, Mexico.
64. Samawi, H. M., Ahmed, M. S. and Abu□Dayyeh. Estimating the Population Mean using Extreme Raked Set Sampling, *The Biometrical Journal*, 38, 577□86.
65. Samawi, H. M. and Muttlak, H. A. (1996). Estimation of Ratio Using Ranked Set Sampling, *Biometrical Journal*, 38, 753□764.
66. Shen, W. H. and Yuan, W. (1996). A Test for a Normal Mean Based on a Modified Partial Ranked Set Sampling, *Pakistan Journal of statistics*, 11, 227□233.
67. Shen, W. H. (1994). Use the Ranked Set Sampling for Testing a Normal Mean, *Calcutta Statistical Association Bulletin*, 44, 183□193.
68. Sinha, B. K., Wu, Z. and Fei, H. (1994). Estimation of a Gamma Mean Based on Ranked Set Sample, *Pakistan Journal of Statistics*, 10, 235□249.
69. Sinha, B. K., A. K. Sinha, and Purkayastha, S. (1996). On some Aspects of Ranked Set Sampling for Estimation of Normal and Exponential Parameters, *Statistics and Decisions*, 14, 223□240.
70. Stokes, S. L. (1977). Ranked Set Sampling with Concomitant Variables, *Commun. Statist. Theory Meth.* A6, 1207□1211.
71. Stokes, S. L. (1980a). Inference on the Correlation coefficient in Bivariate Normal Distribution from Raked Set Samples, *Journal of American Statistical Association*, 75, 989□995.
72. Stokes, S. L. (1980b). Estimation of Variance Using Judgment Ordered Ranked Set Samples, *Biometrics*, 36, 35□42.
73. Stokes, S. L. and Sager, T. W. (1988). Characterization of a Ranked□Set Sample with Application to Estimating Distribution Function, *Journal of American Statistical Association*, 83, 374□381.
74. Stokes, S. L. (1995). Parametric Ranked Set Sampling, *Annals of the Institute of Statistical Mathematics*, 47, 465□482.

75. Takahasi K. and Wakimoto K. (1968). On Unbiased Estimates of the Population Mean Based on the Sample Stratified by Means of Ordering, *Annals of the Institute of Statistical Mathematics*, 21, 249–55.
76. Tam, C. Y. C., Yu, P. L. H. and Fung, T. W. K. (1998). Sensitivity Analysis of BLUE for the Population Mean based on a Ranked Set Sample, *Commun. Statist. Theory Meth.*, 27, 1075–1091.
77. Yu, P. L. H. and Lam, K. (1997). Regression Estimator in Ranked Set Sampling, *Biometrics*, 53, 1070–1080.
78. Yu, P. L. H., Lam, K. and Sinha, B. K. (1997). Estimation of Mean Based on Unbalanced Ranked Set Sample, Applied Statistical Science II, 87–97
79. Yu, P. L. H., Lam, K. and Sinha, B. K. (1999). Estimation of Normal Variance Based on Balanced and Unbalanced Ranked Set Samples, *Environmental and Ecological Statistics*, 6, 23–4

الفصل الحادي والعشرون

طريقة السكين الحادة لإعادة المعاينة
Jackknife Resampling Method

21.1 مقدمة

إن أول من صمم طريقة السكين الحادة (Jackknife) هو Quenouille (1949,1956) لتقليل التحيز للمقدّر، ومن المميزات الجذابة لهذه الطريقة أنه يمكن استخدامها للحالات المعقدة حيث يستحيل تطبيق النماذج المعلماتية أو المعالجات النظرية.

21.2 الطريقة العامة

لقد أعطى Quenouille (1949) وصفاً لتقنية تقليل التحيز لمقدّر الترابط المتسلسل، الذي يعتمد على تقسيم العينة إلى مجموعتين. ثم قام في بحثه المنشور في عام 1956 بتعميم هذه الطريقة على الوجه الآتي: نقوم بتقسيم العينة g من المجاميع المتساوية الحجم وليكن h، أي إن حجم العينة n=gh. لنفرض أن $X_1, X_2, \ldots, X_n$ عبارة عن عينة عشوائية، ولنفترض أن $\hat{\theta}$ عبارة عن المقدّر للمعلمة غير المعلومة θ الذي يعتمد على العينة بحجم n. لنفرض $\hat{\theta}_i$ عبارة عن تقدير θ بالاعتماد على عينة بحجم $(g-1)h$، حيث تم حذف المجموعة i والتي حجمها h.

لنفترض

$$\tilde{\theta}_i = g\hat{\theta} - (g-1)\hat{\theta}_{-i} \quad (i = 1, \ldots, g)$$

لذا فإن المقدّر الآتي يكون تحيزه بالرتبة أو المرتبة n^{-2}

$$\tilde{\theta} = \frac{1}{g}\sum_{i=1}^{g}\tilde{\theta}_i = g\hat{\theta} - \frac{g-1}{g}\sum_{i=1}^{g}\hat{\theta}_{-i}$$

لقد سمى (1958) Tukey المقدّر $\tilde{\theta}$ بالسكين الحادة (jackknife)، التي يمكن أن تكون أداة إحصائية قاسية وجاهزة للاستعمال عند الحاجة.

قام Miller(1974) بمسح جميع النتائج المنشورة حول طريقة السكين الحادة، وهذا البحث يُعَدُّ أساسًا لما تبعه من بحوث في هذا المجال، حيث إن كثيراً من البحوث اللاحقة في هذا المجال تفترض أن g=n و h=1.

لقد اقترح Quenouille(1956) مقدّراً آخر يمكن استخدامه لتقديرات حد $O(n^{-2})$ من التحيز، يسمى مقدّر الدرجة الثانية، وهو

$$\tilde{\theta}^{(2)} = \frac{n^2\tilde{\theta} - (n-1)^2\sum_{j=1}^{n}\tilde{\theta}_{-j}/n}{n^2 - (n-1)^2}$$

حيث إن $\tilde{\theta}_{-j}$ هو عبارة $\tilde{\theta}_i$ ولكن بحجم عينة هو $(n-1)$ حيث تم حذف الوحدة j. ولكن إذا جرى كتابة المقدّر $\tilde{\theta}^{(2)}$ بدلالة المقدّر الأصلي $\hat{\theta}$ فسنحصل على

$$\tilde{\theta}^{(2)} = (2n-1)^{-1}\Big[n^3\hat{\theta} - (2n^2-2n+1)(n-1)\Big(\frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{-i}\Big)$$

$$+ (n-1)^2(n-2)\Big\{\frac{2}{n(n-1)}\sum_{i<j}^{n}\hat{\theta}_{-ij}\Big\}\Big]$$

حيث إن $\hat{\theta}_{-ij}$ عبارة عن المقدّر الأصلي عندما يكون حجم العينة $(n-2)$ حيث جرى حذف المشاهدتين i وj. لذا فإن

$$E(\tilde{\theta}^{(2)}) = \theta + O(n^{-3})$$

لقد اقترح كل من(Gray and Owen (1971 تعديلاً للوزن لكي نحصل على مقدّر غير متحيز عندما يكون التحيز يحتوي على الدرجة الأولى والثانية بدلالة $1/n$ أي

$$E(\hat{\theta}) = q + \frac{a_1}{n} + \frac{a_2}{n^2}$$

لذا فإن المقدّر الذي اقترحاه هو

$$\tilde{\theta}^{(2)^*} = \frac{1}{2}\left[n^2\hat{\theta} - 2(n-1)^2\left(\frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{-i}\right) + (n-2)^2\left\{\frac{2}{n(n-1)}\sum_{i<j}\hat{\theta}_{-ij}\right\}\right]$$

مثال: لنفترض أن θ تمثل متوسط المجتمع و $\hat{\theta} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$. حيث إن $h=1$ و$n=g$ لذا فإن

$$\hat{\theta}_{-i} = (n\hat{\theta} - x_i)/(n-1)$$

لنفترض أن

$$\hat{\theta}_{(.)} = \frac{1}{n}\sum_{i=1}^{n}\hat{\theta}_{-i}$$

لذا فإن

$$\hat{\theta}_{(.)} = \hat{\theta}, \quad \hat{\theta}_{-i} - \hat{\theta}_{(.)} = (\bar{x} - x_i)/(n-1)$$

لذا فإن مقدّر التباين $\hat{\theta}_{(.)} = \hat{\theta}$ هو

$$s^2_{\hat{\theta}_{(.)}} = \frac{n-1}{n}\sum_{i=1}^{n}(\hat{\theta}_{-i} - \hat{\theta}_{(.)})^2 = \frac{1}{n(n-1)}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

وهو نفس المقدّر لتباين متوسط العينة $\bar{X}$.

3.21 التطبيقات الأساسية

يمكـن اسـتخدام طريقـة السـكين الحـادة بشـكل مثمـر في تقـديرات النسـبة والانحـدار. للعينـة ثنائيـة المـتغيرات $(Y_i, X_i), i = 1, 2, ..., n$، حيـث إن $E(X_i) = \mu_x$ و $E(Y_i) = \mu_y$، ونرغب في تقدير $\theta = \mu_y/\mu_x$. في مسوح العينات نفترض أن متوسط المجتمع μ_x للمتغير الإضافي أو الاحتياطي يكون معلوماً، لذا فإن تقدير μ_y هو $\hat{\mu}_y = \hat{\theta}\mu_x$، حيث إن $\hat{\theta} = \overline{Y}/\overline{X}$ وهو ما يسمى تقدير النسبة باستخدام عينة بحجم n. وما هو متعارف عليه أن $\hat{\mu}_y$ أكثر دقة في تقدير μ_y من $\overline{Y}$.

قام Durbin (1959) بتقدير β باستخدام النسبة من خلال طريقة السـكين الحادة بالنموذج

$$Y_i = \alpha + \beta X_i + e_i$$

حيـث إن e_i تتـوزع بشـكل مسـتقل ومتماثـل وتتبـع التوزيـع الطبيعـي أو توزيـع جاما، ولقد قام بدراسة مقدّر السكين الحادة β باستخدام العلاقة

$$\tilde{\theta} = \frac{1}{g}\sum_{i=1}^{g}\tilde{\theta}_i = g\hat{\theta} - \frac{g-1}{g}\sum_{i=1}^{g}\hat{\theta}_{-i}$$

مع g=2. لقد أثبت أن التحيز بالرتبة أو المرتبة n^{-4} ويمكـن إهماله؛ لـذا فـإن مقدّر السكين الحادة لديه تحيز أقل وتباين أقل من $\hat{\theta} = \overline{Y}/\overline{X}$ إذا كانت e_i تتبع التوزيع الطبيعي، أما إذا كانت e_i تتبع توزيع جاما ومع معامل تغير أقل من 1/4، فإن مقدّر السكين الحادة سوف يقلل التحيز، ويزيد التباين قليلاً، لذا سيؤدي إلى خفض متوسط مربعات الخطأ عند مقارنته بالمقدّر $\hat{\theta} = \overline{Y}/\overline{X}$.

ولقد قام كل من Rao (1965) و Rao and Webster (1966) بإعطاء الاختيار الأمثل g و n للتوزيعين الطبيعي وجاما، ولابد من الأخذ في الحسبان أن هناك مقدّرات بديلة بالإضافة إلى مقدّر السكين الحادة، وأن مقدّر السكين

الحادة لا يكون دائماً الأفضل، ولكنه ليس متخلفاً كثيراً عن المقدّر الأمثل، وأخيراً فقد قدم Brillinger (1966) تطبيقات متعددة لاستخدام طريقة السكين الحادة في مسوحات المعاينة المختلفة.

4.21 التقدير بفترة

اقترح Tukey (1958) أن قيم المجموعات التي عددها g وهي $\tilde{\theta}_i \; i=1,\dots,g$ تكون تقريبا متغيرات مستقلة عن بعضها الآخر وتتبع التوزيع نفسه في حالات كثيرة؛ لذا فإن

$$\sqrt{g}(\tilde{\theta}-\theta)/s_{\tilde{\theta}}$$

حيث إن

$$(g-1)s_{\tilde{\theta}}^2 = \sum_{i=1}^{g}(\tilde{\theta}_i-\tilde{\theta})^2$$

تتبع توزيع t التقريبي مع درجات حرية $(g-1)$. يمكن استخدام هذا الإحصاء لإيجاد فترة ثقة θ. ولقد تم إثبات ما ذهب إليه Tukey من أن $\sqrt{g}(\tilde{\theta}-\theta)/s_{\tilde{\theta}}$ تتبع توزيع t أو التوزيع الطبيعي إذا كانت قيمة g كبيرة.

5.21 التحويل

لقد اقترح الباحثون الذين يدافعون عن طريقة السكين الحادة بأن نقوم بمحاولة تثبيت التباين وذلك بالقيام بتحويل المقدّر قبل تطبيق طريقة السكين الحادة. على سبيل المثال يكون أفضل استخدام طريقة السكين الحادة على $\log(s^2)$ و $\tanh^{-1}(r)$ بدلاً من S^2 و r حيث إن r يمثل معامل ارتباط العينة. في بعض الأحيان يكون التحويل واجباً لتفادي تشويه أو تحريف النتائج. على سبيل المثال إذا كنا نرغب في تقدير σ^2، فإنه من دون التحويل ستكون بعض القيم $ns^2-(n-1)s_{-i}^2$ سالبة، وبما أن طريقة السكين الحادة لا ترى القيم

السالبة، لذا لابد من التحويل باستخدام $\log(s^2)$ للتخلص من هذه المشكلة. لمزيد من المعلومات يراجع Govindarajulu(1999).

6.21 التحيز بتقدير التباين

لقد بحث كل من Efron and Stein (1981) تقدير التباين $S(X_1,...,X_n)$ وهي عبارة عن دالة متماثلة للمتغيرات (X_i) العشوائية المستقلة والمتماثلة. لقد برهنا على أن طريقة السكين الحادة لتقدير التباين دائماً تعطينا تقديرات متحيزة نحو الأعلى.

لقد قدم لنا كل من Quenouille and Tukey تقديرات مفيدة للتباين والتحيز ذات الطبيعة غير المعلماتية، نفترض أن $S(X_1,...,X_n)$ عبارة عن مقدّر مصمم لتقدير معلمة ما نرغب في دراستها، نفترض أن S متماثلة $X_i, i=1,...,n$. لنفترض

$$S_{(i)}=(X_1,X_2,...,X_{i-1},X_{i+1},...,X_n)$$

وهي عبارة عن قيمة S تم حسابها بعد حذف X_i. لذا فإن تقدير التباين $var[S(X_1,...,X_n)]$ باستخدام طريقة السكين الحادة هو

$$s^2(S(X_1,...,X_n))=\frac{n-1}{n}\sum_{i=1}^{n}(S_{(i)}-\bar{S})^2$$

حيث إن

$$\bar{S}=\frac{1}{n}\sum_{i=1}^{n}S_{(i)}$$

نلاحظ أن هذا التقدير هو عبارة عن تقدير التباين باستخدام طريقة السكين الحادة S معطى بالشكل $nS-(n-1)\bar{S}$، مع ذلك يمكن أن يكون تقدير التباين S نفسها أو $S_{(i)}$. لمزيد من المعلومات يراجع Efron and Stein (1981) و Govindaragulu(1999).

References

1. Brillinger, D. R. (1966). The Application of Jackknife to the Analysis of Sample Surveys, *Commentary*, 8, 74–80.
2. Booth, J. G. and Hall, P. (1993). An Improvement of the Jackknife Distribution Function Estimator, *Ann. Statist.*, 21, 1476–1485.
3. Durbin, J. (1959). A Note in the Application of Quenouille's Method of Bias Reduction to the Estimation of Ratios, *Biometrika* 46, 477–480.
4. Eforn, B. (1982). The Jackknife, Bootstrap and other Resampling Plans, Siam Publication No. 38, Philadelphia, PA.
5. Eforn, B. (1992). Jackknife-after-Bootstrap Standard Errors and Influence Functions (with Discussions), *J. R. Statist. Soc.*, B54, 83–127.
6. Eforn, B. and Stein, C. (1981). The Jackknife Estimate of Variance, *Ann. Statist.*, 9, 586–596.
7. Frangos, C. C. and Schucany, W. R. (1990). Jackknife Estimation of the Bootstrap Acceleration Constant, *Compu. Statist. Data Anal.*, 9, 271–282.
8. Good, P. H. (2001). Resampling Methods: A Practical Guide to Data Analysis, 2end., Birkhauser, Boston.
9. Govindarajulu, Z. (1999). Elements of Sampling Theory and Methods, Prentice Hall.
10. Gray, H. L. Watkins, T. A. Adams, J. E. (1972). On the Jackknife Statistic, its Extensions, and its Relations to e_n-Transformations, *Ann. Math. Statist.* 43, 1–30.
11. Kunsch, H. R. (1989). The Jackknife and Bootstrap for General Stationary Observations. *Ann. Statist.* 17, 1217–1241.
12. Miller, R. G. (1964). A Trustworthy Jackknife, *Ann. Math. Statist.* 35, 1594–1605.
13. Miller, R. G. (1974). The Jackknife–a Review, *Biometrika* 61, 1–15.
14. Nagao, H. (1988). On the Jackknife Statistics for Eigenvalues and Eigenvectors of a Correlation Matrix, *Ann. Inst. Statist. Math.*, 40, 477–489.
15. Parr. W. C. (1983). A Note on the Jackknife, Bootstrap, and the Delta Method Estimates of Bias and Variance, *Biometrika* 70, 719–722.
16. Parr. W. C. (1985). Jackknifing Differentiable Statistical Functionals, *J. R. Statist. Soc.*, B47, 56–66.
17. Parr. W. C. and Schucany, W. R. (1982). Jackknifing L-Statistics with Smooth Weight Functions, *J. Amer. Statist. Assoc.* 77, 629–638.
18. Quenuouille, M. H. (1949). Approximate Test of Correlation in Time Series, *J. R. Statist. Soc.*, B11, 68–84.
19. Quenuouille, M. H. (1956). Notes on Bias in Estimation, *Biometrika* 52, 647–649.
20. Rao, J. N. K. (1965). A Note on the Estimation of Ratios by Quenouille's Method, *Biometrika* 52, 647–649.
21. Rao, J. N. K. and Shao, J. (1992). Jackknife Variance Estimation with Survey Data under Hot Deck Imputation, *Biometrika* 79, 811–822.
22. Reeds, J. A. (1978). Jackknifing the Maximum Likelihood Estimates, *Ann. Statist.*, 6, 727–739.
23. Schucany, W. R. and Sheather, S. J. (1989). Jackknifing R-Estimators, *Biometrika* 76, 393–398.
24. Rao, J. N. K. and Webster, J. (1966). On Two Methods of Bias Reduction in the Estimation of Ratios, *Biometrika* 53, 571–577.
25. Shao, J. (1988). Consistency of Jackknife Estimators of the Variance of Sample Quantiles, *Comm. Statist.* A17, 3017–3028.
26. Shao, J. (1989). Jackknifing Weighted Least Square Estimators, *J. R. Statist. Soc.*, B51, 139–156.
27. Shao, J. (1992a). One-Step Jackknife for M-Estimators Computed using Newton's Method, *Ann. Inst. Statist. Math.*, 44, 687–701.

28. Shao, J. (1992b). Jackknifing Generalized Linear Models, *Ann. Inst. Statist. Math.*, 44, 673–686.
29. Shao, J. and Tu, D. (1995). The Jackknife and the Bootstrap, Springer, New York.
30. Simonoff, J. S. and Tasi, C. (1988). Jackknifing and Bootstrapping Quasi–Likelihood Estimators, *J. Statist. Compu. Simul.*, 30, 213–232.
31. Stute, W. and Wang, J. (1994). The Jackknife Estimate of a Kaplan–Meire Integral, *Biometrika* 81, 602–606.
32. Tukey, J. W. (1958). Bias and Confidence in Not–Quite Large Samples (Abstract), *Ann. Math. Statist.* 29, 614.
33. Wu, C. F. (1986). Jackknife, Bootstrap and other Resampling Methods in Regression Analysis (with Discussions), *Ann. Statist.*, 14, 1261–1350.

الفصل الثاني والعشرون

طريقة البوتستراب لإعادة المعاينة
Bootstrap Resampling Method

1.22 مقدمة

إن من أهم مغريات استخدام طريقة السكين الحادة والبوتستراب أنه يمكن استخدامهما عندما تكون الحالات التي نرغب في دراستها معقدة ولا تتوافر النماذج المعلماتية ولا المعالجات النظرية. يُعَدُّ (1979) Eforn الرائد لطريقة البوتستراب، ومن ثم توسيع الطريقة لحالات عديدة، على سبيل المثال تقدير وسيط المجتمع حيث أثبت أن البوتستراب أفضل من طريقة السكين الحادة، لمزيد من المعلومات يراجع Eforn and Tibshirani(1993).

2.22 طريقة البوتستراب (Bootstrap Method)

لنفترض أن $X_1, X_2, ..., X_n$ عبارة عن عينة عشوائية من مجتمع لديه $f(x)$ و$f(x)$ دالّتا التوزيع الاحتمالي والكثافة الاحتمالية على التوالي. نفترض أن θ المعلمة التي نرغب في دراستها، ولنفترض أن $\hat{\theta}(X_1, ..., X_n)$ هو تقدير إلى θ والتي تكون متماثلة في $X_1, X_2, ..., X_n$ ، نستطيع أن نكتب الانحراف المعياري إلى $\hat{\theta}$ كما يأتي

$$\sigma_{\hat{\theta}} = \sigma(F, n, \hat{\theta}) = \sigma(F)$$

إن استخدام $\sigma(F)$ في العلاقة أعلاه للتأكيد على أن الانحراف المعياري عبارة عن دالة بدلالة التوزيع الاحتمالي غير المعروف F. إن تقدير البوتستراب إلى الانحراف المعياري يمكن الحصول عليه باستبدال F غير المعروفة بـ $\hat{F}$ وهي عبارة عن تقدير الاحتمالية العظمى إلى F لنحصل على

$$\hat{\sigma}_{\hat{\theta}} = \sigma(\hat{F})$$

لتوضح طريقة البوتستراب لنفترض أننا سحبنا عينة عشوائية بحجم n، متوسط العينة $\bar{X}$ يُعَدُّ تقديراً لمتوسط المجتمع μ. الانحراف المعياري إلى $\bar{X}$ هو

$$(\mu_2/n)^{1/2}$$

حيث إن

$$\mu = E(X) \text{ و } \mu_2 = E(X-\mu)^2$$

لذا فإن

$$\hat{\sigma}_{\bar{x}} = (\hat{\mu}_2/n)^{1/2}$$

حيث إن

$$\hat{\mu}_2 = n^{-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

حيث إن $x_1, \ldots, x_n$ و $\bar{x}$ يمثلون قيم المشاهدات ومتوسطها للمتغيرات $X_1, \ldots, X_n$ و $\bar{X}$ على التوالي. بما أن $\hat{\mu}_2$ متحيز نحو الأسفل أي $E(\hat{\mu}_2)<\mu_2$ ، يمكننا ضرب μ_2 بـ $n(n-1)$ لنحصل على

$$SD = \hat{\sigma}_{\bar{x}} = \{(n-1)^{-1}\sum_{i=1}^{n}(x_i-\bar{x})^2\}^{1/2}$$

في العادة $\sigma(F)$ لا يوجد لها صيغة أو علاقة محددة أو صريحة؛ لذا من أجل حساب قيمة SD لابد من اتباع الخطوات الآتية:

1- لنفترض أن $\hat{F}$ عبارة عن تقدير الاحتمالية العظمى إلى F أي تعطي احتمال $1/n$ لكل مشاهدة x_i.

2- نقوم بسحب عينة البوتستراب من $\hat{F}$ وبالتحديد $X^*_1, ..., X^*_n$ والتي تتوزع بصورة متماثلة كما $\hat{F}$ ومن ثم نحسب $\hat{\theta}^* = \hat{\theta}(X^*_1, ..., X^*_n)$.

3- نقوم بإعادة الخطوة 2 عدة مرات وبالتحديد B مرة (وتكون قيمة B كبيرة) لنجد قيم البوتستراب $\hat{\theta}^*_1, ..., \hat{\theta}^*_B$ ومن ثم نحسب

$$SD = \{ \sum_{j=1}^{B} (\hat{\theta}^*_j - \bar{\hat{\theta}}^*)^2/(B-1) \}^{1/2}$$

حيث إن

$$\bar{\hat{\theta}}^* = \sum_{j=1}^{B} \hat{\theta}^*_j/B$$

إذا افترضنا أن $B \to \infty$ فإننا سنحصل على

$$SD = \{ \sum_{j=1}^{B} (\hat{\theta}^*_j - \bar{\hat{\theta}}^*)^2/(B-1) \}^{1/2} = \sigma_{\hat{\theta}}$$

أي أن القيمة التقديرية باستخدام طريقة البوتستراب ستساوي القيمة التقديرية فيما لو كنا نعرف F. ولكن في الحياة العملية نريد أن تكون قيمة B محدودة بسبب تكاليف الحسابات، بالاعتماد على بعض الحسابات التي قام بها Efron (1982) فإن $B=100$ تعطينا الدقة نفسها بالتقدير فيما لو كانت قيم B تساوي 200 أو 512 أو 1000.

مثال 1: سحبنا عينة عشوائية بحجم 5 من مواليد الماعز ووجدنا أوزانها (1,2,3,4,5) كلغم.

أولاً نحسب

$$\bar{x} = (1+2+3+4+5)/5 = 3$$

و

$$s^2 = \frac{1}{4}\sum_{i=1}^{5}(x_i - 3)^2 = 2.5$$

لذا فإن الانحراف المعياري إلى $\bar{X}$ هو

$$\hat{\sigma}_{\bar{x}} = \frac{s}{\sqrt{n}} = (2.5/5)^{1/2} = 0.707$$

الآن لنطبق طريقة البوتستراب، لنسحب $B=10$ عينة بوتستراب من $(1,2,3,4,5)$ وهي عبارة عن عينة عشوائية مع الإرجاع باستخدام الكمبيوتر أو جداول الأعداد العشوائية، حصلنا على العينات الآتية:

$(1, 2, 2, 3, 5)$ و $(3, 4, 5, 5, 2)$ و $(1, 3, 4, 1, 2)$ و $(5, 3, 1, 4, 2)$ و $(5, 4, 4, 2, 5)$ و $(4, 1, 2, 2, 5)$ و $(3, 3, 4, 1, 5)$ و $(4, 4, 2, 4, 2)$ و $(2, 5, 3, 2, 4)$ و $(5, 4, 1, 5, 5)$

والتي تعطينا

$\hat{\theta}_1^* = 2.6$ و $\hat{\theta}_2^* = 3.8$ و $\hat{\theta}_3^* = 2.2$ و $\hat{\theta}_4^* = 3$ و $\hat{\theta}_5^* = 4$ و $\hat{\theta}_6^* = 3.2$ و $\hat{\theta}_7^* = 2.8$ و $\hat{\theta}_8^* = 3.2$ و $\hat{\theta}_9^* = 3.2$ و $\hat{\theta}_{10}^* = 4$ وأخيراً $\bar{\hat{\theta}}^* = 3.2$.

و

$$\sum_{j=1}^{10}(\hat{\theta}_j^* - \bar{\hat{\theta}}^*)^2 = 3.2$$

لذا فإن تقدير الانحراف المعياري باستخدام طريقة البوتستراب هو

$$SD = (3.2/9)^2 = 0.596$$

نلاحظ أن قيمة $\hat{\sigma}_{\bar{x}} = 0.707$ ليست قريبة من $SD = 0.596$ وهذا يعود لكون قيمة $B=10$ صغيرة جداً.

3.22 طرق البوتستراب للمشكلات العامة

لنفترض أن $\underline{X} = (X_1, ..., X_n)$ و $R(\underline{X}, F)$ عبارة عن دالة إلى $\underline{X}$ ونرغب في دراستها، لنفترض أننا نرغب في تقدير بعض صفات التوزيع الاحتمالي إلى R مثل $E_F(R)$ أو $P_F(R<a)$ لقيمة محددة إلى a. ومن ثم نتبع الخطوات الثلاث السابقة نفسها، ولكن في الخطوة الثانية نحسب

$$R^* = R(\underline{X}^*, \hat{F})$$

بدلاً من $\hat{\theta}^*$. وفي الخطوة الثالثة نحسب صفات R التي نحن بصدد دراستها، على سبيل المثال إذا كنا نرغب في تقدير $E_F(R)$ نحسب

$$E_*(R^*) = B^{-1} \sum_{j=1}^{B} R_j^*$$

و إذا كنا نرغب في تقدير $P_F(R<a)$ نحسب

$$P_*(R^* < a) = \{\#(R_j^* < a)\} / B$$

4.22 تقدير التحيز بالبوتستراب

إذا كنا نرغب في تقدير التحيز لتقدير دالي $\theta(\hat{F})$ إلى $\theta(F)$ وبالتحديد

$$\theta(\hat{F}) - \theta(F)$$

نأخذ

$$R(\underline{X}, F) = \theta(\hat{F}) - \theta(F) = \theta(\hat{F}^*) - \theta(\hat{F}) = \hat{\theta}^* - \hat{\theta}$$

حيث إن $\hat{\theta}^* = \theta(\hat{F}^*)$ و $\hat{F}^*$ عبارة عن التوزيع التجريبي (empirical) الذي يعتمد على عينة البوتستراب $\underline{X}^*$ ، أي أن $\hat{F}^*$ تقوم بإعطاء احتمال مقداره M^*/n لكل x_i ، حيث إن M^* تمثل عدد المرات التي تظهر فيها x_i في عينة البوتستراب.

تقدير البوتستراب للتحيز هو

$$Bias = E_*(R_*) = B^{-1}\sum_{j=1}^{B}\hat{\theta}^*_j - \hat{\theta} = \bar{\hat{\theta}}^* - \hat{\theta}$$

في المثال أعلاه $\bar{\hat{\theta}}^* = 3.2$ و $\hat{\theta} = 3.0$ لذا فإن التحيز

$$Bias = \bar{\hat{\theta}}^* - \hat{\theta} = 3.2 - 3 = 0.2$$

لمزيد من المعلومات يراجع Eforn (1982, 1992) أو Govindarajulu(1999).

5.22 مشكلة الانحدار

لقد تناولنا إلى الآن طريقة البوتستراب لعينة واحدة أو ما يسمى مشكلة العينة الواحدة التي يكون فيها المتغير العشوائي X_i لديه نفس التوزيع الاحتماليF. ولكن طرق البوتستراب يمكن تطبيقها إلى حالات أكثر تعقيداً، فيما يلي سوف نستخدم طرق البوتستراب على مشكلة الانحدار.

لنأخذ النموذج الآتي:

$$Y_i = g_i(\beta) + \varepsilon_i \quad i = 1,\dots,n.$$

حيث y_i تمثل المشاهدات للمتغير $Y_i(i = 1,\dots,n)$. نفترض الدوال $g_i(.)$ معلومة الشكل وتعتمد على موجه معين ثابت c_i. بينما β_{px1} موجه للمعلمات غير المعروفة، أما حدود الخطأ ε_i فتتوزع بشكل مستقل ومتماثل كما F، حيث $E_F(\varepsilon_i) = 0$ أو الوسيط إلى ε_i يساوي صفر. بعد مشاهدة موجه المتغير

$\underline{Y}=(Y_1,\dots,Y_n)'$ ولـــيكن $\underline{y}=(y_1,\dots,y_n)'$ نرغـــب في تقـــدير الموجـــه β باستخدام أحد المعايير مثل تصغير $D(\underline{y},\eta)$ المسافة بين $\underline{y}$ وموجه المتنبئات

$$\eta(\beta)=(g_1(\beta),\dots,g_n(\beta))'$$

لذا فإن

$$\hat{\beta}=\min_{\beta} D(\underline{y},\eta(\beta))$$

إن الاختيار المناسب إلى D هو

$$D(\underline{y},\eta)=\sum_{i=1}^{n}(y_i-\eta_i)^2$$

حيث إن $\eta_i=g_i(\beta)$.

نفترض أن النموذج فيه من التعقيد ما يجعل تحليله بالطرق المعروفة صعباً جـــداً، علـــى ســـبيل المثـــال $g_i(\beta)=\exp(\underline{c}_i\beta)$ و F غـــير معروفـــة و $D(\underline{y},\eta)=\sum_{i=1}^{n}|y_i-\eta_i|$. لـذا فـإن الخطـوات الـثلاث الـسابقة يمكـن تعديلـها لتصبح

1- ابنِ $\hat{F}$ التي تعطي احتمال $1/n$ لكل مشاهدة من مشاهدات المتبقي $\hat{\varepsilon}_i$ والذي يعرف

$$\hat{\varepsilon}_i=y_i-g_i(\hat{\beta})$$

2- اسحب عينة البوتستراب من $\hat{F}$ وبالتحديد $\varepsilon^*_1,\dots,\varepsilon^*_n$ ، لذا فإن عينة البوتستراب تعرف

$$Y^*_i=g_i(\hat{\beta})+\varepsilon^*_i,\quad i=1,\dots,n.$$

حيث إن ε^*_i تتوزع بشكل مستقل ومتماثل كما $\hat{F}$ ، ومن ثم نحسب

$$\hat{\beta}^*=\min_{\beta} D(\underline{Y}^*,\eta(\beta))$$

3- نقوم بإعادة الخطوة الثانية B مرة (وتكون قيمة B كبيرة) لنجد قيم البوتستراب $\hat{\beta}_1^*, ..., \hat{\beta}_B^*$.

نفترض أننا نرغب في تقدير مصفوفة التغاير إلى $\hat{\beta}$ ، لذا فإن تقدير البوتستراب إلى مصفوفة التغاير هو

$$Cov = \frac{1}{B-1}\sum_{j=1}^{B}(\hat{\beta}_j^* - \bar{\hat{\beta}}^*)(\hat{\beta}_j^* - \bar{\hat{\beta}}^*)'$$

مثال 2: لنفترض أن Y تمثل عمر الشخص وc دخل الشخص بآلاف الدولارات. ونفترض أن نموذج الانحدار هو

$$Y_i = \beta c_i + \varepsilon_i, i = 1, ..., n$$

حيث إن $E(\varepsilon_i) = 0$ و $var(\varepsilon_i) = \sigma^2$ ، وقد حصلنا على البيانات الآتية:

$$(20, 0.8), (25, 1.1), (30, 1.2), (35, 1.3), (40, 1.5)$$

نرغب في تقدير تباين $\hat{\beta}$. نقدر β باستخدام طريقة المربعات الصغرى

$$\hat{\beta} = \sum_{i=1}^{5} c_i y_i / \sum_{i=1}^{5} c_i^2 = 185/4750 = 0.039$$

$$\hat{y}_i = \hat{\beta} c_i$$

$$\hat{\varepsilon}_i = y_i - \hat{y}_i = y_i - \hat{\beta} c_i$$

لذا فإن

$$\hat{\varepsilon}_1 = 0.8 - 0.78 = 0.02, \hat{\varepsilon}_2 = 1.1 - 0.975 = 0.125, \hat{\varepsilon}_3 = 0.03, \hat{\varepsilon}_4 = -0.065, \hat{\varepsilon}_5 = -0.06$$

أما تباين $\hat{\beta}$ فيمكن حسابه باستخدام

$$var(\hat{\beta}) = \frac{\hat{\sigma}^2}{\sum c_i^2}$$

حيث

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{5} \hat{\varepsilon}_i}{5-1} = \frac{0.02475}{4} = 0.0062$$

لذا فإن تباين $\hat{\beta}$ هو

$$\text{var}(\hat{\beta}) = \frac{\hat{\sigma}^2}{\sum c_i^2} = \frac{0.0062}{7.23} = 0.0008575$$

لنستخدم طريقة البوتستراب لتقدير تباين $\hat{\beta}$.سنطبق المثال بسحب عشر عينات للبوتستراب أي $B=10$ ، أولاً نسحب عشر عينات حجم كل عينة 5 وحدات من الأرقام العشوائية بين 1 و5 ولنفترض أن الأرقام التي حصلنا عليها في المثال الأول تمثل هذه العينات، على سبيل المثال العينة الأولى (1,2,2,3,5) والعينة الثانية (3,4,5,5,2) وهكذا؛ لذا فإن عينة البوتستراب الأولى للمتبقي هي: $\varepsilon_1^* = \hat{\varepsilon}_1 = 0.02$ و $\varepsilon_2^* = \hat{\varepsilon}_2 = 1.1$ و $\varepsilon_3^* = \hat{\varepsilon}_2 = 1.1$ و $\varepsilon_4^* = \hat{\varepsilon}_3 = 0.03$ و أخيراً $\varepsilon_5^* = \hat{\varepsilon}_5 = -0.06$. وهكذا لجميع العينات. الآن نستخدم $\varepsilon_i^*, i = 1, ..., 5$ للعينة الأولى لحساب Y_i^* باستخدام

$$Y_i^* = \hat{\beta} c_i + \varepsilon_i^* \quad i = 1, ..., n.$$

الآن لحساب Y_1^* نقوم بالتعويض عن قيم ε_1^* في

$$Y_1^* = 0.039 c_1 + \varepsilon_1^* = 0.039(20) + 0.02 = 0.8$$

لذا فإن الزوج الأول من قيم عينة البوتستراب هو: $(Y_1^*, c_1) = (0.8, 20)$.

لتوضيح الفكرة نحسب (Y_3^*, c_3)

$$Y_3^* = 0.039 c_3 + \varepsilon_3^* = 0.039(20) + 1.1 = 1.295$$

لــــذا فـــإن الـــزوج الثالـــث مـــن قـــيم عينـــة البوتســـتراب هـــو: $(Y_3^*, c_3) = (1.295, 30)$. وهكذا لجميع قيم العينة لنجد:

$$(0.8, 20), (1.1, 25), (1.295, 30), (1.395, 35), (1.56, 40)$$

نستخدم هذه العينة لحساب قيمة β^* للعينة الأولى باستخدام طريقة المربعات الصغرى لنحصل على $\hat{\beta}^* = 0.0408$.

الجدول الآتي يعطينا قيم جميع عينات البوتستراب العشر وقيمة $\hat{\beta}^*$ لكل عينة.

	العينة	$\hat{\beta}^*$
1	(0.8, 20), (1.1 , 25), (1.295, 30), (1.395, 35), (1.56, 40)	0.0408
2	(0.81 , 20), (0.91 , 25), (1.1 , 30), (1.305, 35), (1.685, 40)	0.0390
3	(0.8, 20), (1.005, 25), (1.1 05, 30), (1.385, 35), (1.685, 40)	0.0400
4	(0.72, 20), (0.978, 25), (1.1 9, 30), (1.365, 35), (1.685, 40)	0.0399
5	(0.72, 20), (0.91 , 25), (1.1 05, 30), (1..49, 35), (1.56, 40)	0.0389
6	(0.81 , 20), (1.005, 25), (1.1 05, 30), (1.385, 35), (1.56, 40)	0.0390
7	(0.71 5, 20), (0.995, 25), (1.295, 30), (1.49, 35), (1.56, 40)	0.0405
8	(0.71 5, 20), (0.91 , 25), (1.295, 30), (1.365, 35), (1.685, 40)	0.0409
9	(0.905, 20), (0.91 5, 25), (1.1 7, 30), (1.49, 35), (1.495, 40)	0.0396
10	(0.72, 20), (0.91 1 , 25), (1.1 9, 30), (1.305, 35), (1.56, 40)	0.0381

المتوسط $\bar{\hat{\beta}}^* = 0.03967$ وتقدير التباين إلى $\hat{\beta}$ باستخدام البوتستراب هو

$$\frac{1}{B-1}\sum_{j=1}^{B}(\hat{\beta}_j^* - \bar{\hat{\beta}}^*)^2 = \frac{1}{9}(0.01574469 - 0.01573709) = 0.00000084$$

نلاحظ أن تقدير التباين باستخدام البوتستراب أقل بكثير من التقدير الحقيقي.

6.22 التقدير بفترة

إذا كنا نرغب في حساب 95% فترة ثقة إلى المعلمة θ يمكننا القيام بذلك على الوجه الآتي:

$$\hat{\theta} \pm 2\,SD$$

ولكن إذا أردنا أن نكون أكثر دقة، يمكننا أن نحسب فترة ثقة بعد إيجاد قيم البوتستراب إلى $\hat{\theta}^*_1, \ldots, \hat{\theta}^*_B$ ، يمكننا إيجاد $(1-\alpha)$ فترة ثقة والتي هي

$$(\theta^*_{k_1}, \theta^*_{k_2})$$

حيث

$$k_1 = [n\alpha/2] \text{ و } k_2 = [n(1-\alpha/2)].$$

إذا استخدمنا البيانات الواردة في المثال أعلاه وبثقة 80%، فسنحصل على فترة ثقة إلى β وهي (0.0381 , 0.0408).

References

1. Abramovitch, L. and Singh, K (1985). Edgeworth Corrected Pivotal Statistics and the Bootstrap, *Ann. Statist.*, 13, 116–132.
2. Arcones, M. A. and Gine, E. (1991). Some Bootstrap Tests of Symmetry for Univariate Continuous Distributions, *Ann. Statist.*, 19, 1496–1551.
3. Babu, G. J. and Bose, A. (1989). Bootstrap Confidence Intervals, *Statist. Prob. Letters*, 7, 151–160.
4. Babu, G. J. and Singh K. (1983). Inference on Means using Bootstrap, *Ann. Statist.*, 11, 999–1003.
5. Bickel, P. J. and Freedman, D. A.(1984). Asymptotic Normality and the Bootstrap in Stratified Sampling, *Ann. Statist.*, 12, 470–482.
6. Boos, D. D. and Brownie, C. (1989). Bootstrap Methods foe Testing Homogeneity of Variances, *Technometrics*, 31, 69–82.
7. Booth, J. G. and Hall, P. (1994). Monte–Carlo Approximation and the Iterated Bootstrap, *Biometrika*, 81, 331–340.
8. Bose, A. (1990). Bootstrap in Moving Average Models, *Ann. Inst. Statist. Math.*, 42, 753–768.
9. Chao, M. T. and Lo, S. H. (1985). A Bootstrap Method for Finite Populations, *Sankhya A*, 47, 399–405.
10. Chen, Z. and Do, K. A. (1992). Important Resampling for the Smoothed Bootstrap, *J. Statist. Compu. Simul.* 40, 107–124.
11. Davison, A. C. and Hall, P. (1993). On Studentizing and Blocking Methods for Implementing the Bootstrap with Dependent Data, *Austral. J. Statist.*, 35, 215–224.
12. Diaconis, P. and Eforn, B. (1983). Computer–Intensive Methods in Statistics, *Scientific American*, May 1983, 116–130.
13. DiCiccio, T. J. and Tibshirani, R. J. (1987).Bootstrap Confidence Intervals and Bootstrap Approximations, *J. Amer. Statist. Assoc.*, 82, 163–170.
14. Do, K. A. and Hall, P. (1991). On Importance Resampling for the Bootstrap, *Biometrika*, 78, 161–167.
15. Eforn, B. (1981). The Jackknife Estimation of Variance, *Ann. Statist.*, 9, 586–596.
16. Eforn, B. (1982). The Jackknife, Bootstrap and other Resampling Plans, Siam Publication No. 38, Philadelphia, PA.
17. Eforn, B. (1992). Jackknife–after–Bootstrap Standard Errors and Influence Functions (with Discussions), *J. R. Statist. Soc.*, B54, 83–127.
18. Eforn, B. and Tibshirani, R. (1993). Introduction to the Bootstrap, Chapman and Hall, New York.
19. Falk, M. (1992). Bootstrap Optimal Bandwidth Selection for Kernel Density Estimates, *J. Statist. Plan. Inference*, 30, 13–32.
20. Freedman, D. A. and Peters, S. C. (1984). Bootstrapping a Regression Equation– Some Empirical Results, *J. Amer. Statist. Assoc.*, 79, 97–106.
21. Frangos, C. C. and Schucany, W. R. (1990). Jackknife Estimation of the Bootstrap Acceleration Constant, *Compu. Statist. Data Anal.*, 9, 271–282.

22. Franklin, L. A. and Wesserman, G. S. (1992). Bootstrap Lower Confidence Limits for Capability Indices, *Journal of Quality Technology*, 24, 196–210.
23. Ghosh, M. Parr, W. C., Singh, K. and Babu, G. J. (1984). A Note on Bootstrapping the Sample Median, *Ann. Statist.*, 12, 1130–1135.
24. Good, P. H. (2001). Resampling Methods: A Practical Guide to Data Analysis, 2end., Birkhauser, Boston.
25. Govindarajulu, Z. (1999). Elements of Sampling Theory and Methods, Prentice Hall.
26. Hall, P. (1986). On the Bootstrap and Confidence Intervals, *Ann. Statist.*, 14, 1431–1452.
27. Hall, P. (1987). On the Bootstrap and Likelihood-Based Confidence Regions, *Biometrika*, 74, 481–493.
28. Hall, P. (1988). Theoretical Comparisons of Bootstrap and Confidence Intervals (with Discussions), *Ann. Statist.*, 16, 927–953.
29. Hall, P. (1989). On efficient Bootstrap Simulation, *Biometrika*, 76, 613–617.
30. Hall, P. (1990). Using the Bootstrap to Estimate Mean Square Error and Select Smoothing Parameters in Nonparametric Problems, *J. Multivariate Anal.*, 32, 177–203.
31. Hall, P. (1992). On Bootstrap Confidence Intervals in Nonparametric Regression, *Ann. Statist*, 20, 689–711.
32. Hall, P. and Martin, M. A. (1988). On the Bootstrap and Two Sample Problems, *Austral. J. Statist.*, 30, 179–182.
33. Hall, P. and Wilson, S. R. (1991). Two Guide Lines for Bootstrap Hypothesis Testing, *Biometrics*, 47, 757–762.
34. Hinkely, D. V. (1987). Bootstrap Significance Tests, *Proceeding of the 47th Session of International Statistical Institute*, 65–74, Paris.
35. Janas, D. (1993). Bootstrap Procedures for Time Series, Shaker, Aachen.
36. Knight, K. (1989). On the Bootstrap of the Sample Mean in the Infinite Variance Case, *Ann. Statist*, 17, 1168–1175.
37. Kuk, A. Y. C. (1987). Bootstrap Estimators of Variance under Sampling with Proportional to Aggregate Size, *J. Statist. Compu. Simul.* 28, 303–311.
38. Kuk, A. Y. C. (1989). Double Bootstrap Estimation of Variance under Systematic Sampling with Probability Proportional to Size, *J. Statist. Compu. Simul.* 31, 303–311
39. Kunsch, H. R. (1989). The Jackknife and Bootstrap for General Stationary Observations. *Ann. Statist*, 17, 1217–1241.
40. Lahiri, S. N. (1991). Second Order Optimality of Stationary Bootstrap, *Statist. Prob. Letters*, 14, 335–341.
41. Lee, K. W. (1990). Bootstrapping Logistic Regression Models with Random Regression, *Comm. Statist. A*, 19, 2527–2539.
42. Liu, R. Y. (1988). Bootstrap Procedures under some non-i.i.d. Models, *Ann. Statist*, 16, 1697–1708.
43. Lo, A. Y. (1987). A Large Sample Study of the Bayesian Bootstrap, *Ann. Statist*, 15, 360–375.
44. Loh, W. Y. (1991). Bootstrap Calibration for Confidence Construction and Selection, *Statist. Sinica*, 1, 479–495
45. Parr. W. C. (1983). A Note on the Jackknife, Bootstrap, and the Delta Method Estimates of Bias and Variance, *Biometrika* 70, 719–722.

46. Politis, D. N. and Romano, J. P. (1994). The Stationary Bootstrap, *J. Amer. Statist. Assoc.*, 89, 1303–1313.
47. Romano, J. P. (1988). Bootstrapping the Mode, *Ann. Inst. Statist. Math.*, 40, 565–586.
48. Shao, J. (1988). Bootstrap Variance and Bias Estimation in Linear Models, *Canadian, J. Statist.*, 16, 371–382.
49. Shao, J. (1992). Bootstrap Variance and Bias Estimators with Truncation, *Statist. Prob. Letters*, 15, 95–101.
50. Shao, J. (1994). Bootstrap Sample Size in Nonlinear Cases, *Proceeding of the Amer. Math. Soc*,. 122, 1251–1262.
51. Shao, J. and Tu, D. (1995). The Jackknife and the Bootstrap, Springer, New York.
52. Sitter, R. R. (1992). Comparing Three Bootstrap Methods for Survey Data, *Canadian, J. Statist.*, 20, 135–154.
53. Stute, W. (1990). Bootstrap of the Linear Correlation Model, *Statistics*, 21, 433–436.
54. Tibshirani, R. J. (1988). Variance Stabilization and the Bootstrap, *Biometrika*, 75, 433–444.
55. Tu, D. and Zhang, L. (1992). On the Estimation of Skewness of a Statistic using the Jackknife and the Bootstrap, *Statistical Papers*, 33, 39–56.
56. Wang, S. (1989). On the Bootstrap and Smoothed Bootstrap, *Comm. Statist. A*, 18, 3949–3962.
57. Wu, C. F. (1986). Jackknife, Bootstrap and other Resampling Methods in Regression Analysis (with Discussions). *Ann. Statist.*, 14, 1261–1350.
58. Yang, S. S. (1988). A Central Limit Theorem for the Bootstrap Mean, *Amer. Statist.*, 42, 202–203.
59. Zhang, J. Boos, D. D. (1992). Bootstrap Critical Values for Testing Homogeneity of Covariance Matrices, *J. Amer. Statist. Assoc.*, 87, 425–429.

الفصل الثالث والعشرون

أخطاء عدم الإجابة
Nonresponse Errors

23.1 مقدمة

بالإضافة إلى أخطاء المعاينة التي هي جزء من عملية المسح بالعينة والتي لا يمكن التخلص منها بشكل نهائي ولكن يمكن تقليلها، هناك أخطاء أخرى يمكن أن تحدث، منها:

1. **أخطاء التقارير.** دَخلُ الشخص الذي يخبر به دائرة الضرائب ربما يختلف عن دخله الحقيقي، كثيرٌ من الناس يعرفون وزنهم وطولهم ومقدار ما يصرفونه شهرياً على أنفسهم ولكن نادراً ما نحصل منهم على المعلومات الدقيقة حول هذه الأمور.
2. **أخطاء عدم الإجابة.** غالباً ما يقوم بعض الأشخاص الذين ترسل لهم الاستبانات برميها في سلة المهملات، ولا يجيبون عليها على الرغم من التذكيرات التي ترسل لهم لاحقاً، وفي بعض الأحيان الذين يجيبون ربما يعودون إلى طبقة معينة، والذين لا يستجيبون يعودون إلى طبقة أخرى، وهذا يؤدي إلى التحيز في العينات.
3. **أخطاء باختيار العينة.** إن اختيار العينة ربما يكون متحيزاً بسبب المسح أو عندما لا يتم اتباع المعاينة العشوائية بشكل صحيح وصارم. على سبيل المثال الفواكه المعروضة في نوافذ المحلات ليست كالتي هي موجودة داخل المحل.

4. **أخطاء عدم قياس بعض الوحدات.** أحياناً لا يقوم العادون أو الباحثون بقياس بعض الوحدات التي هي من ضمن وحدات العينة، وهذا يحدث عندما تكون عملية العد في الليل، كذلك يحدث في المجتمعات الإنسانية وذلك بسبب عدم الاستطاعة من تحديد أماكن بعض الوحدات.
5. **أخطاء في قياس الوحدة.** ربما تكون وحدة القياس متحيزة أو غير دقيقة. في المجتمعات الإنسانية ربما لا توجد لدى الأشخاص المعلومات الدقيقة، أو ربما يقومون بإعطاء أجوبة متحيزة.
6. **أخطاء فنية.** تحدث هذه الأخطاء في عملية معالجة البيانات مثل عملية الترميز أو التحرير أو جدولة البيانات أي وضعها أو تبويبها في جداول.

2.23 تأثير عدم الإجابة

لنمعن النظر في تأثير عدم الإجابة على تقدير العينة للمتوسط والنسبة. لنفترض أن N_1 يمثل عدد المستجيبين إلى استبانة معينة (لنرمز لها بالطبقة I). لنفترض أن $N = N_1 + N_2$ حيث إن $W_2 = N_2/N$ و $W_1 = N_1/N$. ذلك يعني أن W_2 تشير إلى نسبة عدم الإجابة في المجتمع. إذا سحبنا عينة عشوائية بحجم n من الطبقة I و $\bar{y}_1$ يمثل تقدير العينة، لذا فإن التحيز في متوسط العينة بسبب عدم الإجابة هو

$$E(\bar{y}_1) - \bar{Y} = \bar{Y}_1 - \bar{Y} = \bar{Y}_1 - (W_1\bar{Y}_1 + W_2\bar{Y}_2) = W_2(\bar{Y}_1 - \bar{Y}_2)$$

بما أنه لا يمكن مشاهدة الطبقة الثانيةII، لذا فإن حجم التحيز غير معروف فيما عدا إذا قمنا بتخمين $\bar{Y}_2$. ومع ذلك إذا كانت البيانات مستمرة فإن فترة الثقة للمتوسط $\bar{Y}_2$ تكون واسعة جداً لدرجة لا يمكن الاستفادة منها. ولكن إذا كنا نرغب في تقدير النسبة P_2 فإن هناك أمل أن تكون

نسبة عدم الإجابة P_2 محدودة من الأعلى بواحد عندما نقوم بحساب فترة ثقة إلى نسبة المجتمع P. لنفترض أن حجم العينة العشوائية الكلية هو n وهنالك n_1 استجابة من بين n وحدة. إذا كان حجم n_1 كبير يمكننا أن نجد فترة 95% ثقة إلى P_1 وهي

$$p_1 \pm 2\sqrt{p_1(1-p_1)/n_1}$$

عندما يمكن إهمال معامل التصحيح للمجتمعات المحدودة، حيث P_1 يمثل نسبة العينة. يمكننا أن نجد فترة ثقة متحفظة إلى نسبة المجتمع P وذلك بوضع قيمة $P_2=0$ عندما نجد الحد الأدنى لفترة الثقة $\hat{P}_L$ و $P_2=1$ عندما نجد الحد الأعلى لفترة الثقة $\hat{P}_U$، وعليه فإن فترة 95% ثقة إلى P ستكون

$$\hat{P}_L = W_1[p_1 - 2\sqrt{p_1(1-p_1)/n_1}\,] + 0(W_2)$$

$$\hat{P}_u = W_1[p_1 + 2\sqrt{p_1(1-p_1)/n_1}\,] + 1(W_2)$$

عندما تكون قيمة W_2 غير معلومة فإن Cochran (1977) اقترح الطريقة الآتية لإيجاد فترة ثقة إلى P. عند حساب $\hat{P}_L$ نفترض أن جميع عدم المستجيبين في العينة أعطوا جواباً سالباً، وهذا يعني أن $W_1=1$ و $P_2=0$. أما عندما نقوم بحساب $\hat{P}_U$ فنفترض أن جميع عدم المستجيبين أعطوا جواباً موجباً، وهذا يعني أن $W_1=1$ و $P_2=1$. لاحظ عندما تكون $W_1=1$ سنضع $n_1=n$.

مثال: لنفترض أن $n=500$ و $n_1=400$ و $p_1=15\%$ ، لذا فإن 75 عضواً من العينة أجابوا بالإيجاب، ومعدل عدم الإجابة هو 20% . أوجد 95% فترة ثقة إلى P.

الحل:

$$\hat{P}_L = 0.15 - 2\sqrt{0.15(0.85)/500} = 0.15 - 0.032 = 11.8\%$$

$$\hat{P}_U = 0.35 + 2\sqrt{0.35(0.65)/500} = 0.35 + 0.043 = 39.3\%$$

3.23 أنواع عدم الإجابة

يمكننا أن نعطي تصنيفاً تقريبياً لأنواع عدم الإجابة وهي

1. **عدم التغطية**. وهو عبارة عن الفشل في تحديد بعض وحدات العينة أو الوصول إليها، ويبرز هذا غالباً عندما يفشل العاد بالوصول إلى بعض الوحدات بسبب رداءة الطقس، أو أن تكون القائمة والكشف الذي عنده غير كامل، أو بسبب عدم توافر المواصلات الجيدة في أثناء مدة المسح ليتسنى الوصول إلى جميع الوحدات.
2. **غير موجود في المنزل**. تتكون هذه الفئة من مجموعة من الأشخاص الذين يقيمون في المنزل ولكنهم غير موجودين في المنزل ساعة حضور العاد إليه. غالباً ما يصعب الحصول على شخص داخل المنزل إذا كان جميع أفراد العائلة يعملون، بعكس العوائل التي فيها أطفال أو عجزة حيث غالباً ما تجد شخصاً ما داخل المنزل.
3. **غير قادر على الإجابة**. قد لا يمتلك الشخص المسؤول الإجابة، التي تتطلب معلومات محددة أو ربما لا يرغب في إعطاء الإجابة، تتولى عملية حسن صياغة الأسئلة في الاستبانة معالجة مثل هذه الأمور.
4. **الفريق الصعب**. الأشخاص الذين يرفضون الإجابة باستمرار أو العاجزين عن إجراء المقابلة أو الذين غالباً ما يكونون غير موجودين في المنزل طيلة مدة المسح هم الذين يشكلون هذا القطاع. وهم يشكلون مصدراً للتحيز لا يتزحزح مهما بذلت الجهود.

إن معالجة مشكلة عدم التغطية غالباً ما تكون صعبة، ولكن هناك بعض الطرق التي يمكن أن نتبعها للتغلب أو التخفيف من هذه المشكلة منها: إعادة زيارة الوحدات التي لم يكن بالمستطاع الوصول إليها في المحاولة الأولى، القيام بوضع القوائم بعناية بحيث يمكن استخدامها للتحقق من أن المعلومات

كاملة. وأحياناً نقوم بالمقارنة بين أعداد الأشخاص أو المنازل التي أحصيناها وبين المعلومات المتوافرة حول الموضوع نفسه من مسوحات أخرى. وعندما يكون الدليل غير كامل وذلك بسبب إنشاء مبانٍ جديدة يتم تدعيم الكشف الموجود بعينة مساحية، الهدف منها معاينة أجزاء من المدينة للتعرف على المباني الجديدة وإضافتها إلى الكشف الموجود، لمزيد من المعلومات حول المسوحات التي تعاني من مشكلة عدم التغطية وكيفية معالجة هذه المشكلة يرجى الرجوع إلى Kish and Hess (1958) وWoolesy (1956).

فيما يتعلق بمشكلة الغائبين عن المنزل تكون المشكلة أسهل إذا كان أي بالغ في المنزل يمكنه الإجابة على الأسئلة، كما هو الحال في المسوحات التي نريد أن نقابل بها شخصاً واحداً محدداً جرى سحبه بطريقة عشوائية. ونفضل مقابلة شخصٍ بالغٍ بمفرده، إذا كان المسح من النوع الذي يتضمن أشخاصاً لا يستطيع أحدهم الإجابة بدقة نيابة عن الآخرين، أو إذا كان هناك ارتباطات عالية ما بين المعلومات ضمن المنزل الواحد، بحيث يصبح قياس أكثر من شخص واحد غير مُجدٍ اقتصادياً. لقد طور Kish (1949) طريقة جيدة لاختيار شخص واحد من أسرة واحدة.

4.23 الرجوع مرة أخرى

لتفادي الأعداد الكبيرة من عدم الإجابة نقوم بالرجوع للوحدة أكثر من مرة، ومن الأشياء المتعارف عليها هو الرجوع مرة أخرى خصوصاً إذا لم يكن هناك أحد في المنزل. لنمعن النظر في الجدول الآتي الذي يتضمن نسبة الإجابة مصحوبة بنسبة المنازل التي لديها أطفال بعمر أقل من سنتين موزعة على عدد مرات الرجوع، هذه البيانات تم الحصول عليها من Hilgard and Payne (1944)

التي رجع إليها P.S.R.S. Rao(1983) والتي تمثل مسحاً يتألف من3265 منزل، والتي نفذت من خلال مقابلة شخصية لأصحاب المنازل.

عدد مرات الرجوع

	1	2	>2	عدم الإجابة	المجموع
نسبة الإجابة	63.5	22.2	14.3	0	10.0
نسبة المنازل التي لديها أطفال صغار من بين المستجيبين	17.2	9.5	6.2	0	13.9

يوضح هذا الجدول أن الشخص البالغ الذي لديه أطفال صغار يكون احتمالية وجوده في المنزل للمقابلة أكبر عندما يقوم العاد بزيارتهم. وهذا يعني إذا توقف العاد بعد الزيارة الأولى فإن العينة تمثل العوائل التي لديها أطفال صغار بنسبة أكبر من العوائل الأخرى. فإذا كان المجتمع الذي نستهدفه هو جميع العوائل فإن ذلك سيؤدي إلى تحيز في تقدير جميع المتغيرات التي لها علاقة بالأطفال الصغار. نتوقف عن إعادة الزيارة أو الرجوع في الحياة العملية بعد محاولات قليلة. والسبب الرئيس للتوقف هو التكاليف، والتي تلعب الدور الكبير في كثيرٍ من المسوحات. ربما نحتاج إلى نموذج لتحديد مجموع تكاليف العمليات الميدانية للمسح بالعينة، يراجع Govindarajulu (1999) وBirnbaum and Sirken (1950) لمزيد من المعلومات.

5.23 أخطاء القياس

تحدث أخطاء القياس في الحياة العملية، ولقد دُرست هذه المسألة من قبل أكثر من باحث من بينهم Sukhatme et al. (1984). لنفترض أن y_i تمثل القيمة الحقيقية للإجابة من قبل الوحدة i ونرمز لها بـ U_i. لنفترض أننا سحبنا عينة عشوائية بسيطة بحجم n وحدة من مجتمع بحجم N وحدة، وأيضًا

نفترض أن x_{ij} تمثل القيمة التي قدمها العاد j للوحدة U_i حيث إن $i=1,\dots,h$ و$j=1,\dots,m$.

لندرس أبسط النماذج وهو

$$x_{ij}=y_i+\alpha_j+\varepsilon_{ij}$$

حيث

α_j : يمثل التحيز للعاد j للمشاهدات المعادة لجميع الوحدات

ε_{ij} : يمثل الخطأ الذي ارتكبه العاد j في الوحدة i

حيث إن

$$E(\varepsilon_{ij}|i,\ j)=0$$

و

$$E(\varepsilon_{ij}^2)=S_e^2$$

و ε_{ij} غير مرتبطة لجميع قيم i وj.

نفترض أن عدد التكرارات n_{ij} تساوي 1 أو 0. لنفترض

$n_{i.}=\sum_{j=1}^{m}n_{ij}$: عدد المشاهدات للوحدة U_i

$n_{.j}=\sum_{i=1}^{h}n_{ij}$: عدد المشاهدات التي قام بها العاد j

$\bar{x}_{.j}=\frac{1}{n_{.j}}\sum_{i=1}^{h}n_{ij}x_{ij}$: المتوسط لجميع المشاهدات التي قام بها العاد j والتي عددها $n_{.j}$

$\bar{x}=\frac{1}{n}\sum_{i}^{h}\sum_{j}^{m}n_{ij}x_{ij}$: المتوسط لجميع المشاهدات n والتي أجريت على h من الوحدات في العينة.

لذا

$$\bar{x}_{.j} = \frac{1}{n_{.j}}\sum_{i=1}^{h}(n_{ij}y_i + n_{ij}\alpha_j + n_{ij}\varepsilon_{ij}) = \frac{1}{n_{.j}}\sum_{i}^{h} y_i n_{ij} + \alpha_j + \frac{1}{n_{.j}}\sum_{i}^{h} n_{ij}\varepsilon_{ij}$$

كذلك

$$\bar{x}_{..} = \frac{1}{n}\sum_{i=1}^{h} y_i n_{i.} + \frac{1}{n}\sum_{j=1}^{m}\alpha_j n_{.j} + \sum_{i}^{h}\sum_{j}^{m}\varepsilon_{ij} n_{ij}$$

لنفترض أننا وضعنا الفروض الآتية:

1. العادون m يمثلون عينة عشوائية بسيطة من مجتمع حجمه M من العادين.
2. الوحدات h في العينة جرى توزيعها بصورة عشوائية بين العادين.
3. $n_{.j} = n/m = \bar{n}$.
4. $n_{i.} = n/h = p.$

لذا فإن المعادلتين أعلاه تصبحان

$$\bar{x}_{.j} = \frac{1}{\bar{n}}\sum_{i=1}^{h} y_i n_{ij} + \alpha_j + \frac{1}{\bar{n}}\sum_{i}^{h} n_{ij}\varepsilon_{ij}$$

و

$$\bar{x}_{..} = \frac{1}{h}\sum_{i=1}^{h} y_i + \frac{1}{m}\sum_{j=1}^{m}\alpha_j + \frac{1}{n}\sum_{i}^{h}\sum_{j}^{m}\varepsilon_{ij} n_{ij}$$

كذلك

$$E(\bar{x}_{.j}) = \frac{1}{N}\sum_{i=1}^{N} y_i + \frac{1}{M}\sum_{i}^{M}\alpha_j = \mu + \bar{\alpha}$$

و

$$E(\bar{x}_{..}) = \mu + \bar{\alpha}$$

يمكننا أن نجد تباين $\bar{x}_{.j}$

$$var(\bar{x}_{.j}) = (\frac{1}{\bar{n}} - \frac{1}{N})S_y^2 + (1 - \frac{1}{M})S_\alpha^2 + \frac{S_\varepsilon^2}{\bar{n}}$$

إذا كانت M وN كبيرتين، فيمكن كتابة العلاقة الأخيرة كما يأتي

$$var(\overline{x}_{.j}) \quad (\frac{1}{n})(S_y^2+S_\varepsilon^2)+S_\alpha^2$$

و تباين $\overline{x}_{..}$

$$var(\overline{x}_{..}) = (\frac{1}{h}-\frac{1}{N})S_y^2+(\frac{1}{m}-\frac{1}{M})S_\alpha^2+\frac{S_\varepsilon^2}{n} \quad \frac{S_y^2}{h}+\frac{S_\alpha^2}{m}+\frac{S_\varepsilon^2}{n}$$

حيث

$$S_\alpha^2 = \frac{1}{M-1}\sum_{j=1}^{M}(\alpha_j-\overline{\alpha})^2 \text{ و } S_y^2=\frac{1}{N-1}\sum_{i=1}^{N}(y_i-m)^2$$

نلاحظ أنه بالرغم من كون α تختلف من عاد إلى آخر لكننا لا نستطيع إهمال $\overline{\alpha}$.

بما أن تباين المشاهدة الواحدة التي سحبت من مجتمع غير متناهٍ عندما تكون M كبيرة يمكننا أن نكتب

$$S_x^2 = S_y^2+S_\alpha^2+S_\varepsilon^2$$

يمكننا أن نكتب تباين $\overline{x}_{..}$

$$var(\overline{x}_{..}) = \frac{S_x^2}{h}+(\frac{1}{m}-\frac{1}{h})S_\alpha^2+(\frac{1}{n}-\frac{1}{h})S_\varepsilon^2$$

نلاحظ من المعادلة الأخيرة أن التباين للمقدر $\overline{x}_{..}$ ليس مصدره الاختلاف بين وحدات العينة، ولكن تم تضخيمه بسبب التغاير أو الاختلاف بين العادين. لذا فإن الصيغ التي مرت معنا في الفصول السابقة قد تُقدر التباين بأقل مما يجب عليه. وأخيراً نلاحظ أن

$$var(\overline{x}_{..}) = \frac{S_x^2}{h}$$

إذا كان m=h=n أو إذا كانت α_j ثابتة لجميع قيم j وh=n. لمزيد من المعلومات يراجع Sukhatme et al. (1984).

6.23 تأثير التحيز الثابت

لنفترض أن هناك تحيزاً ثابتاً مقداره β في مقاسات المتغير y_i لجميع الوحدات ومقداره غير معلوم. لذا فإن متوسط العينة العشوائية البسيطة $\bar{y}$ سيكون عرضة لخطأ مقداره β. ولكن التحيز يلغى عند تقدير خطأ التباين على اعتبار أنه يعتمد على مجموع المربعات $(y_i - \bar{y})^2$. وكنتيجة لذلك فإن حساب فترة الثقة لمتوسط المجتمع لن تخضع لهذا التحيز، وسنحصل على النتيجة نفسها إذا استخدمنا العينة الطبقية.

وتبقى الحالة نفسها بالنسبة لتقدير النسبة والانحدار. لنأخذ تقدير الانحدار

$$\bar{y}_{lr} = \bar{y} + b(\bar{X} - \bar{x})$$

إذا كان y_i و x_i يخضع كل منهما إلى تحيز ثابت مقداره β_y و β_x على التوالي. بما أن تقدير المربعات الصغرى إلى b لم يتغير بسبب التحيز والتحيز يلغى من الحد $(\bar{X} - \bar{x})$و هذا يعني أن $\bar{y}_{lr}$ سيتأثر بتحيز β_y. ويمكن بسهولة إثبات أن تقدير تباين $\bar{y}_{lr}$ لا يحتوي على أي تأثير للتحيز.

أما فيما يخص تقدير النسبة

$$\bar{y}_R = \frac{\bar{y}}{\bar{x}}\bar{X}$$

فإن التحيز الثابت سيكون مقدارهβ_y ، بما أن $E(\bar{X}|\bar{x}) \approx 1$ إذا كان حجم العينة كبيراً حتى ولو كان x_i يتعرض إلى تحيز ثابت. في حالة العينات الكبيرة فإن تقدير التباين هو

$$s^2(\bar{y}_R) = \frac{(N-n)}{Nn}\frac{\sum(y_i - \hat{R}x_i)^2}{n-1}$$

سيكون حراً من التحيز باعتباره تقديراً إلى

$$E(\bar{y}_R - \bar{Y})^2$$

خلاصة القول أن التحيز الثابت سوف يمر من غير أن تكتشفه بيانات الإحصائية للعينة.

7.23 الإجابة العشوائية

عندما يطلب من الشخص الإجابة عن بعض الأسئلة الحساسة مثل: الإصابة بمرض فقد المناعة المكتسبة أو لديه علاقة غير شرعية، غالباً ما يتهرب الشخص من الإجابة عن مثل هذا النوع من الأسئلة أو لا يعطي الجواب الصحيح. في مثل هذه الحالات ربما نلجأ إلى طريقة الإجابة العشوائية، على سبيل المثال إذا كانت π_A تمثل النسبة الحقيقية من الناس الذين يحملون الصفة A، لقد أثبت Warner (1965) أنه يمكن تقدير π_A من دون أن يحتاج الشخص المجيب أن يكشف نفسه فيما يخص هذا السؤال الحساس أو المحرج.

نختار الإجابة العشوائية مثل رمي زهر النرد، صندوق يحتوي على كرات حمراء وبيضاء، اختيار عبارة من عبارتين كل منهما يتطلب الإجابة عليها بنعم أو لا، كي يجري تقديمه إلى المستجيب، ولا يعلم العاد أو المعاين أي سؤال أجاب عليه المستجيب. ولكن العاد يعلم الاحتمالات P و$1-P$ التي قدمت بها العبارتان. الفكرة الأساسية هنا أن المستجيب متأكد أنه لم يفشِ سره حول السؤال الحساس.

لنفترض أن العبارتين أو الجملتين كما يأتي

" أنا أحمل الصفةA " (وتقدم باحتمالP).

" أنا لا أحمل الصفةA " (وتقدم باحتمال$1-P$).

في عينة عشوائية حجمها n، نفترض أن لدينا m، جواب إيجابي أي أنا أحمل الصفة A. لذا فإن

$$\hat{\phi} = m/n$$

حيث

$$\phi = P\pi_A + (1-P)(1-\pi_A) = (2P-1)\pi_A + 1 - P$$

عندما تكون P معلومة

$$\hat{\pi}_A = \frac{\hat{\phi} - (1-P)}{2P-1}, \quad P \neq 1/2$$

يمكننا أن نلاحظ أن $\hat{\pi}_A$ عبارة عن تقدير الاحتمالية العظمى (MLE) إلى π_A و $E(\hat{\pi}_A) = \pi_A$ وبما أن $\hat{\phi}$ هو تقدير الاحتمالية العظمى إلى ϕ و $E(\hat{\phi}) = \phi$ وتباين $\hat{\pi}_A$ هو

$$var(\hat{\pi}_A) = \frac{\phi(1-\phi)}{n(2P-1)^2}$$

باستخدام قيمة ϕ أعلاه يمكننا إعادة كتابة تباين $\hat{\pi}_A$ كما يأتي:

$$var(\hat{\pi}_A) = \frac{\pi_A(1-\pi_A)}{n} + \frac{P(1-P)}{n(2P-1)^2}$$

سيكون القسم الأول من تباين $\hat{\pi}_A$ أعلاه التباين الفعلي لو أن جميع الأشخاص أجابوا على السؤال الحساس بصورة صحيحة فيما إذا كان يحملون الصفة A. ممكن أن يكون القسم الثاني كبيراً جداً، وهذا يعتمد على قيمة P.

مثال: ترغب إحدى المدارس أن تقدر نسبة الطلاب الذين يغشون في الاختبارات من بين طلاب المدرسة. قام العاد بسحب عينة عشوائية حجمها

n=500 طالب من طلاب المدرسة التي عددها N=2500. كل طالب سئل أن يسحب ورقة من بين مجموعة أوراق، في هذه المجموعة 85% من الأوراق تحمل العبارة " أنا أغش في الاختبارات" و15% تحمل العبارة " أنا لا أغش في الاختبارات". ولقد أجاب 300 طالب بنعم للورقة التي رآها، علماً بأن العاد لا يعرف ما كتب على الورقة التي سحبها الطالب وأجاب عليها. أوجد فترة 95% ثقة لنسبة الطلاب الذين يغشون من بين طلاب المدرسة.

الحل:

$$\hat{\phi}=\frac{m}{n}=\frac{300}{500}=0.6$$

$$\hat{\pi}_A=\frac{\hat{\phi}-(1-P)}{2P-1}=\frac{0.6-(1-0.85)}{2(0.85)-1}=0.643$$

$$\text{var}(\hat{\pi}_A)=\frac{\pi_A(1-\pi_A)}{n}+\frac{P(1-P)}{n(2P-1)^2}=\frac{0.64(1-0.64)}{500}+\frac{0.85(1-0.85)}{500(2(0.85)-1)^2}$$

$$=0.0004591+0.0005204=0.0009795$$

لذا فإن فترة 95% ثقة لنسبة الغشاشين في المدرسة هي

$$0.643\pm 2\,(0.0313)=0.643\pm 0.0626$$

لمزيد من المعلومات ولتفاصيل أكثر يراجع Warner (1965) وCochran (1977) وGovindarajulu (1999).

8.23 اختيار إعادة المقابلة

لقد لاحظنا أعلاه أن هناك نسبة من عدم الاستجابة، سببها في الغالب عدم وجود الشخص في وقت المقابلة أو الذهاب إليه لجمع المعلومات منه في

وقت غير مناسب. إذا كان لدينا عينة حجمها n من الوحدات وحصلنا على n_1 من الإجابات حيث إن $n_1 < n$ ، سيكون لدينا مجموعة من عدم الإجابة ، يمكن أن نتصور أن المجتمع مكون من طبقتين، عدد وحدات العينة التي اختيرت من الطبقة الأولى n_1 ومن الطبقة الثانية $n_2 = n - n_1$ ، ولكن في الحقيقة ليس لدينا طبقية بالمعنى الصحيح، لأن حجم العينة من الطبقة الثانية n_2 متغير عشوائي يمكن تحديده فقط بعد الانتهاء من جمع البيانات، ولكن إذا فكرنا فيها كعينة طبقية نستطيع أن نحصل على طريقة مثالية لتقسيم الموارد المالية بين المسح الأولي والإعادة.

لنفترض أننا قررنا أن نعيد مقابلة r من الذين لم نستطع مقابلتهم في المحاولة الأولى والبالغ عددهم n_2 حيث إن $r = n_2/k$ و $k > 1$ عدد ثابت.

إذا كان $\bar{y}_1$ يمثل معدل العينة في المحاولة الأولى و $\bar{y}_2$ معدل العينة إلى r من الوحدات في المحاولة الثانية، إذاً

$$\bar{y}^* = \frac{1}{n}(n_1 \bar{y}_1 + n_2 \bar{y}_2)$$

تقدير غير متحيز إلى الوسط الحسابي للمجتمع، لمزيد من المعلومات يراجع Scheaffer, Mendenhall and ott (1996).

References

1. Armitage, P. (1947). A Comparison of Stratified with Unrestricted Random Sampling from a finite Population. *Biometrika*, 34, 273–280.
2. Bartholomew, D. J. (1961). A Method for "not-at-home" Bias in Sample Surveys, *App. Stat.* 10, 52–59.
3. Birnbaum, Z., W. and Sirken, M. G. (1950). Basis Due to Nonavailability In Sampling Surveys, *J. Amer. Statist. Assoc.*, 45, 98–111.
4. Cochran, W. G. (1977). Sampling Technique, 3rd Ed. Wiley, New York.
5. Deming, W. E. (1953). On a Probability Mechanism to Attain an Economic Balance Between the Resultant Error of Non-Response and the Bias of Non-Response, *J. Amer. Statist. Assoc.*, 48, 743–772.
6. Durbin, J. (1954). Non-Response and Call-Backs in Surveys, *Bull. Int. Stat. Inst.*, 34, 72–86.
7. Durbin, J. and Stuart, A. (1954). Callbacks and Clustering in Sample Surveys: An Experimental Study, *J. Roy. Stat. Soc.* A117, 387–428.
8. Govindarajulu, Z. (1999). Elements of Sampling Theory and Methods, Prentice Hall.
9. Hansen, M. H. (1951). Response Errors in Surveys, *J. Amer. Statist. Assoc.*, 46, 147–190.
10. Hansen, M. H. and Hurwitz, W. N. (1946). The Problem of Nonresponse in Sample Surveys, *J. Amer. Statist. Assoc.*, 41, 517–529.
11. Hansen, M. H. and Waksberg, J. (1970). Research on Non-Response in Censuses and Surveys, *Rev. Int. Stat. Inst.*, 38, 318–332.
12. Hoewitz, D. G., Shah, B. V. and Simmons, W. R. (1967). The Unrelated Nonresponse Randomized Response Model, *Poroc. Soc. Statist. Section Amer. Statist. Assoc.*, 663–685.
13. Kish, L (1949). A Procedure for Objective Respondent Selection within the Household, *J. Amer. Statist. Assoc.*, 44, 380–387.
14. Kish, L and Hess, I (1958). On Noncoverage of Sample Dwelling, *J. Amer. Statist. Assoc.*, 53, 509–524.
15. Kish, L. and Lansing, J, B. (1954). Response Errors in Estimating the Value of Homes, *J. Amer. Statist. Assoc.*, 49, 520–538.
16. Madow, G. A. (1965). On some Aspects of Response Error Measurement, *Poroc. Soc. Statist. Section Amer. Statist. Assoc.*, 182–192.
17. Oh, H. L. and Scheuren, F. J. (1983). Weighting Adjustment for Unit Nonresponse. In *Incomplete Data in Sample Surves,* Vol. 2, (ed. I. Olkin and D. B. Rubin), New York, Academic Press, 143–184.
18. Politz, A. N. and Simmons, W. R. (1949). An Attempt to Get the "Not at Homes" into the Sample without Callbacks, *J. Amer. Statist. Assoc.*, 44, 9–31.
19. Politz, A. N. and Simmons, W. R. (1950). An Attempt to Get the "Not at Homes" into the Sample without Callbacks, *J. Amer. Statist. Assoc.*, 45, 136–137.
20. Rao, P. S. R. S. (1983). Callbacks, Followups and Repeated Telephone Calls. In *Incomplete Data in Sample Surves,* Vol. 2, (ed. I. Olkin and D. B. Rubin), New York, Academic Press, 33–44.
21. Scheaffer, R.L., Mendenhall, W. and Ott L. (1996). Elementary Sampling Survey, 5th ed., Duxbury, New York.

22. Sukhatme, P. V. and Seth, G. R. (1952). Non-Sampling Errors in Surveys, *J. Ind. Soc. Agr. Stat.*, 4, 5-41.
23. Sukhatme, P. V. and Sukhatme, B. V. (1984). Sampling Theory of Surveys with Applications, (3rd ed.) Ames, Iowa State Univ. Press.
24. Warner, S. L. (1965). Randomized Response-A Survey Technique for Eliminating Evasive Answer Bias, *J. Amer. Statist. Assoc.*, 60, 63-69.
25. Woolsey, T. D. (1956). Sampling Methods for Small Household Survey, *Pub. Health Monograph*, No. 40.

الفصل الرابع والعشرون

طرق أخرى للمعاينة
Other Sampling Methods

1.24 مقدمة

إن الهدف من هذا الفصل إعطاء القارئ فكرة عن بعض طرق المعاينة الأخرى، التي قد يحتاجها الباحث في عمله أو في أثناء قيامه بالبحث العلمي. سوف نتناول باختصار شديد بعض طرق المعاينة التي قد تستعمل أكثر من غيرها في الحياة العملية، وسوف يعطى القارئ مراجع لكل طريقة سنتكلم عنها في هذا الفصل للرجوع إليها ودراستها بشكل مفصل.

2.24 المعاينة العشوائية مع الإرجاع

لقد تكلمنا بالتفصيل عن العينة العشوائية البسيطة بدون إرجاع. وكذلك غالبية طرق المعاينة التي تكلمنا عنها في هذا الكتاب هي بدون إرجاع بشكل أو آخر. وذلك لأهمية وانتشار استخدامها في الحياة العملية.

إن الفكرة الأساسية لعملية المعاينة مع الإرجاع كما أشرنا إلى ذلك في الفصل الأول تقوم على سحب وحدات العينة من المجتمع ، باستخدام إحدى الطرق العشوائية المعروفة ولكن نقوم بإرجاع الوحدة التي تم سحبها إلى المجتمع بعد مشاهدتها قبل أن نقوم بسحب وحدة جديدة وهكذا. تسمى هذه الطريقة (Random Sampling with Replacement) المعاينة العشوائية مع الإرجاع. قد يكون سبب عدم انتشار أو استعمال هذه الطريقة في الحياة العملية على الرغم من سهولة التعامل مع التقديرات الإحصائية باستخدامها ،

أنَّ الوحدة إذا تم سحبها أكثر من مرة في العينة فإنها في المرات اللاحقة لن تضيف أي معلومات جديدة، لأنها مكررة، لمزيد من المعلومات يراجع Cochran (1977) أو ترجمة أنيس كنجو(1995).

3.24 التقدير في المجتمعات الجزئية

نرغب في بعض الأحيان التقدير لمجموعة جزئية من المجتمع أو لمجتمع جزئي، بالإضافة إلى المجتمع كاملاً، فمثلاً قد نرغب في تقدير دخل النساء في المجتمع بالإضافة إلى تقدير دخل الفرد في المجتمع، تشكل النساء هنا مجموعة جزئية من المجتمع الكلي الذي يحتوي على الرجال والنساء، أو قد نرغب في تقدير دخل العاملين في قطاع الزراعة بالإضافة إلى دخل جميع أفراد المجتمع، لمزيد من المعلومات يراجع(1977) Cochran أو ترجمة أنيس كنجو (1995) وكذلك (1972) Raj.

4.24 نموذج الإجابة العشوائية

غالباً ما يرفض كثير من الأشخاص الذين نجمع المعلومات منهم الإجابة عن بعض الأسئلة التي قد تسبب لهم في بعض الأحيان أذى، فعلى سبيل المثال غالباً يرفض الأشخاص الإجابة عن الأسئلة السياسية في البلدان التي تكون فيها الديمقراطية معطلة، لأن الإجابة عن هذه الأسئلة قد تسبب لهم أذى.

سوف نحاول تقدير نسبة الأشخاص الذين يحملون صفة معينة دون الحاجة إلى مساءلتهم بطريقة مباشرة في المقابلة. إن أول من اقترح هذه الطريقة هو(1965) Warneer، فمثلاً إذا أردنا تقدير نسبة الطلاب الذين يحاولون الغش في الامتحانات لن نحصل على إجابات صحيحة إذا سألنا الطلاب بصورة مباشرة، لذلك نقوم بتقسيم المجتمع إلى مجموعتين A وB في مثالنا هذا قد تكون A يغش وB لا يغش، وأعضاء المجتمع هنا إما يقعون في A أو B.

لنفترض أن p تمثل نسبة الطلاب الذين يغشون في الامتحان. إن الهدف هو تقدير p دون أن نسأل الطلاب مباشرة في المقابلة.

نبدأ بأخذ مجموعة من الأوراق البيضاء التي كلها تحمل نفس المواصفات، ما عدا جزءاً ونسبة θ يحمل علامة A وجزءاً آخر ونسبة $(1-\theta)$ يحمل علامة B. نقوم بسحب عينة عشوائية بسيطة حجمها n طالب، ونطلب من كل واحد منهم سحب ورقة واحدة، ويجيب بنعم إذا كان الحرف المكتوب على الورقة يتفق مع مجموعته، ولا، إذا كان يختلف عن مجموعته. إن الشخص الذي يقوم بالمقابلة لا يرى ما هو مكتوب على الورقة، فقط يسجل الإجابات نعم أو لا وبعد الانتهاء من جميع أفراد العينة، نستخدم هذه البيانات لتقدير p، يراجع Scheaffer, Mendenhall and ott (1996) لمزيد من المعلومات.

5.24 المعاينة العنقودية بثلاث مراحل

إن طريقة العينة العنقودية بمرحلتين يمكن أن توسع إلى مرحلة ثالثة، وذلك بسحب عينة عشوائية من كل وحدة من وحدات المرحلة الثانية بدلاً من مشاهدة جميع عناصر الوحدة، على سبيل المثال في مسح الحقول لتقدير إنتاجية أحد المحاصيل كالقمح مثلاً، يمكن أن نَعُدّ المرحلة الأولى هي سحب عينة من القرى الموجودة في المنطقة التي يجري فيها تقدير المحصول، أما المرحلة الثانية فتكون اختيار مجموعة من الحقول المزروعة بالقمح الموجودة في القرى التي سحبت في المرحلة الأولى، أما المرحلة الثالثة فتكون اختيار عدد من القطع الصغيرة داخل كل حقل سحب في المرحلة الثانية. نلاحظ أننا حصلنا على العينة العنقودية في ثلاث مراحل وذلك من خلال توسيع المرحلة الثانية في العينة العنقودية بمرحلتين، مع الافتراض أن جميع الوحدات في المرحلة الثالثة نختارها باحتمالات متساوية. لمزيد من المعلومات يراجع Singh and Chaudhary (1986).

6.24 الطبقية مع المعاينة العنقودية

إن من أهم الطرق الشائعة الاستعمال على مستوى كبير في المسوحات ما يسمى (Cluster Sampling with Stratification) الطبقية مع العينة العنقودية. لا توجد هنالك مفاهيم جديدة مع هذه الطريقة. عندما نريد تقدير الوسط الحسابي للمجتمع على سبيل المثال نقوم بتقسيم المجتمع إلى k من الطبقات، وتكون المعاينة ضمن كل طبقة مستقلة عن الطبقة الأخرى، حيث نقوم بسحب عينة عنقودية داخل كل طبقة.

لنفترض أن الطبقة i تحتوي N_i من الوحدات (المرحلة الأولى) وكل وحدة تحتوي على M_i من الوحدات (المرحلة الثانية) لذلك فإن n_i هو حجم العينة المسحوبة في المرحلة الأولى وm_i هو حجم العينة المسحوبة في المرحلة الثانية إن تقدير الوسط الحسابي للمجتمع.

حيث إن $\overline{y}_i$ يمثل الوسط الحسابي للطبقة i و

$$W_i = \frac{N_i M_i}{\sum_{i=1}^{k} N_i M_i}$$

يمثل وزن الطبقة لمزيد من المعلومات يراجع Singh and Chaudhary (1986).

7.24 المعاينة بالحصة

تؤدي السرعة عملاً مهماً في مسوحات رأي الناس، لذلك أصبح شائعاً استخدام ما يسمى بالمعاينة بالحصة (Quota Sampling) لاختبار العينة من المجتمع. إن طريقة المعاينة بالحصة قائمة على مبدأ أن العينة توزع جيداً على المجتمع، ويجب أن تحتوي على نفس النسبة التي يمثلها أشخاص يحملون صفة مميزة في المجتمع، فمثلاً إذا كانت نسبة الرجال في المجتمع 60% فيجب أن تحتوي العينة على 60% من الرجال. إن أهم الصفات التي في العادة تؤخذ في

الحسبان في هذا النوع من المعاينة هي الجنس، والعمر، والوظيفة، والحالة الاقتصادية، وحجم السكان بالإضافة إلى التوزيع الجغرافي للمجتمع.

إن المعاينة بالحصة تستخدم المسوحات الشاملة أو عينات كبيرة سحبت من المجتمع في مُدَدٍ سابقة متقاربة. لتقسيم المجتمع إلى طبقات والحصول على أوزان هذه الطبقات، يقوم العادون بسحب أشخاص من المجتمع بأي طريقة تعجبهم على شرط أن يحافظوا على الحصة المحددة لكل طبقة، أي مثلاً يجب أن يكون عدد الرجال في العينة 60% والنساء 40% أو أي صفة تستخدم كالعمر أو الوظيفة ...إلخ.

إن الفرق الرئيس بين العينة بالحصة والعينة العشوائية البسيطة أو الطبقية هو أن العينة بالحصة لا تستخدم العشوائية في اختيار الوحدات من المجتمع، لذلك ربما يقرر العادُّ إهمال أجزاء من المجتمع إذا كان لا يستطيع الوصول إليها بسهولة. لذلك من الناحية النظرية البحتة يمكن أن نقول إن العينة بالحصة تفتقر إلى الأساس العلمي، لأنه لا يدخل فيها عنصر العشوائية في الاختيار، أي اختيار وحدات العينة من المجتمع، لذلك لا نستطيع استخدام نتائجها دون الخوف من أن هنالك تحيزاً في اختيار وحدات العينة، ولكنها تستخدم كثيراً في أبحاث السوق واستطلاعات الرأي، وأحياناً تعطي نتائج قريبة من العينة الاحتمالية.

إن من أهم ميزات المعاينة بالحصة أنها سريعة، وتقلل التكاليف والجهد والوقت. لمزيد من المعلومات يراجع Cochran (1977) أو ترجمة أنيس كنجو (1995) و(1991) Barnett و(1968) Raj.

المراجع العربية

1. وليم كوكوران – ترجمة أنيس كنجو ،(1995) تقنية المعاينة الإحصائية، جامعة الملك سعود، الرياض.

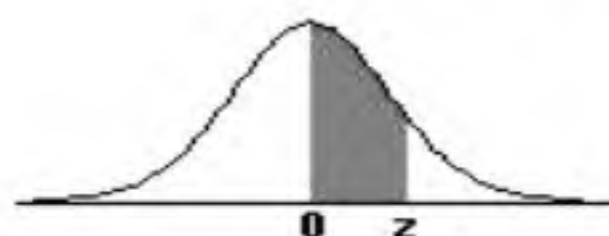
References

1. Barnett, V. (1991). Sample Survey Principles and Methods, Oxford Uni. Press, New York.
2. Chaudhuri, A. and Stenger, H. (1992). Survey Sampling: Theory and Methods, Marcel Dekker, New York.
3. Cochran, W. G. (1977). Sampling Technique, 3rd Ed. Wiley, New York.
4. Deming, W. E. (1960). Sampling Design in Business Research. Wiley, New York
5. Foreman, E. K. (1991). Survey Sampling Principles, Marcel Dekker, New York.
6. Govindarajulu, Z. (1999). Elements of Sampling Theory and Methods, Prentice Hall.
7. Hajek, J (1981). Sampling from Finite Population, Marcel Dekker, New York.
8. Hansen, M. H., and Hurwitz, W. N. (1943). On the Theory of Sampling from Finite Populations. *Ann. Math. Statist.* 14, 333–362.
9. Hedayat, A. S. and Sinha, B. K. (1991). Design and Inference in Finite Population Sampling, Wiley, New York.
10. Kish, L (1995). Survey Sampling, Wiley, New York.
11. Levy, P. S. and Lemeshow S. (1991). Sampling of Population: Methods and Applications, Wiley, New York.
12. Lohr, S. L. (1999). Sampling: Design and Analysis, Duxbury, New York.
13. Raj, D. (1968). Sampling Theory, McGraw Hill, New York.
14. Sampath, S. (2000). Sampling Theory and Methods, CRC Press.
15. Scheaffer, R.L., Mendenhall, W. and Ott L. (1996). Elementary Sampling Survey, 5th ed., Duxbury, New York.
16. Singh, D. and Chaudhary, F. S. (1986). Theory and Analysis of Sample Survey Designs, Wily Eastern, New Delhi.
17. Sturat A. (1984). The Ideas of Sampling, Revised Edition, Griffin, London.
18. Sukhatme, P. V., Sukhatme, B. V., Sukhatme, S., and Ashok, C. (1984). Sampling Theory of Surveys with Applications, 3rd ed., Ames (Iowa): Iowa State University Press.
19. Thompson, S. K (2002). Sampling, 2end Wiley, New York.
20. Warner, S. L. (1965). Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias, *J. Amer. Statist. Assoc.*, 60, 63–69.

الملاحق

أ. جدول التوزيع الطبيعي المعياري (Z) وجدول توزيع t

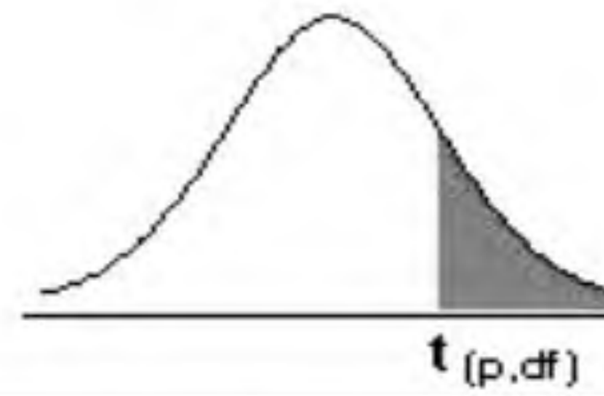
أ. جدول التوزيع الطبيعي المعياري [1]



z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

(1) www.statsoftinc.com/textbook/sttable.html

ب . جدول توزيع t



α / ν	0.4	0.25	0.1	0.05	0.025	0.01	0.005	0.0005
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657	636.62
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925	31.599
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841	12.924
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604	8.610
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032	6.869
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707	5.959
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499	5.408
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355	5.041
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250	4.781
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169	4.587
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106	4.437
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055	4.318
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012	4.221
14	0.258	0.692	1.345	1.761	2.145	2.624	2.977	4.141
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947	4.073
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921	4.015
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898	3.965
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878	3.922
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861	3.883
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845	3.850
21	0.257	0.686	1.323	1.721	2.080	2.518	2.831	3.819
22	0.256	0.686	1.321	1.717	2.074	2.508	2.819	3.792
23	0.256	0.685	1.319	1.714	2.069	2.500	2.807	3.768
24	0.256	0.685	1.318	1.711	2.064	2.492	2.797	3.745
25	0.256	0.684	1.316	1.708	2.060	2.485	2.787	3.725
26	0.256	0.684	1.315	1.706	2.056	2.479	2.779	3.707
27	0.256	0.684	1.314	1.703	2.052	2.473	2.771	3.690
28	0.256	0.683	1.313	1.701	2.048	2.467	2.763	3.674
29	0.256	0.683	1.311	1.699	2.045	2.462	2.756	3.659
30	0.256	0.683	1.310	1.697	2.042	2.457	2.750	3.646
∞	0.253	0.674	1.282	1.645	1.960	2.326	2.576	3.291

2. الحقيبة الإحصائية MINTTAB

الجزء الأول

سحب عينة عشوائية باستخدام الأعداد العشوائية

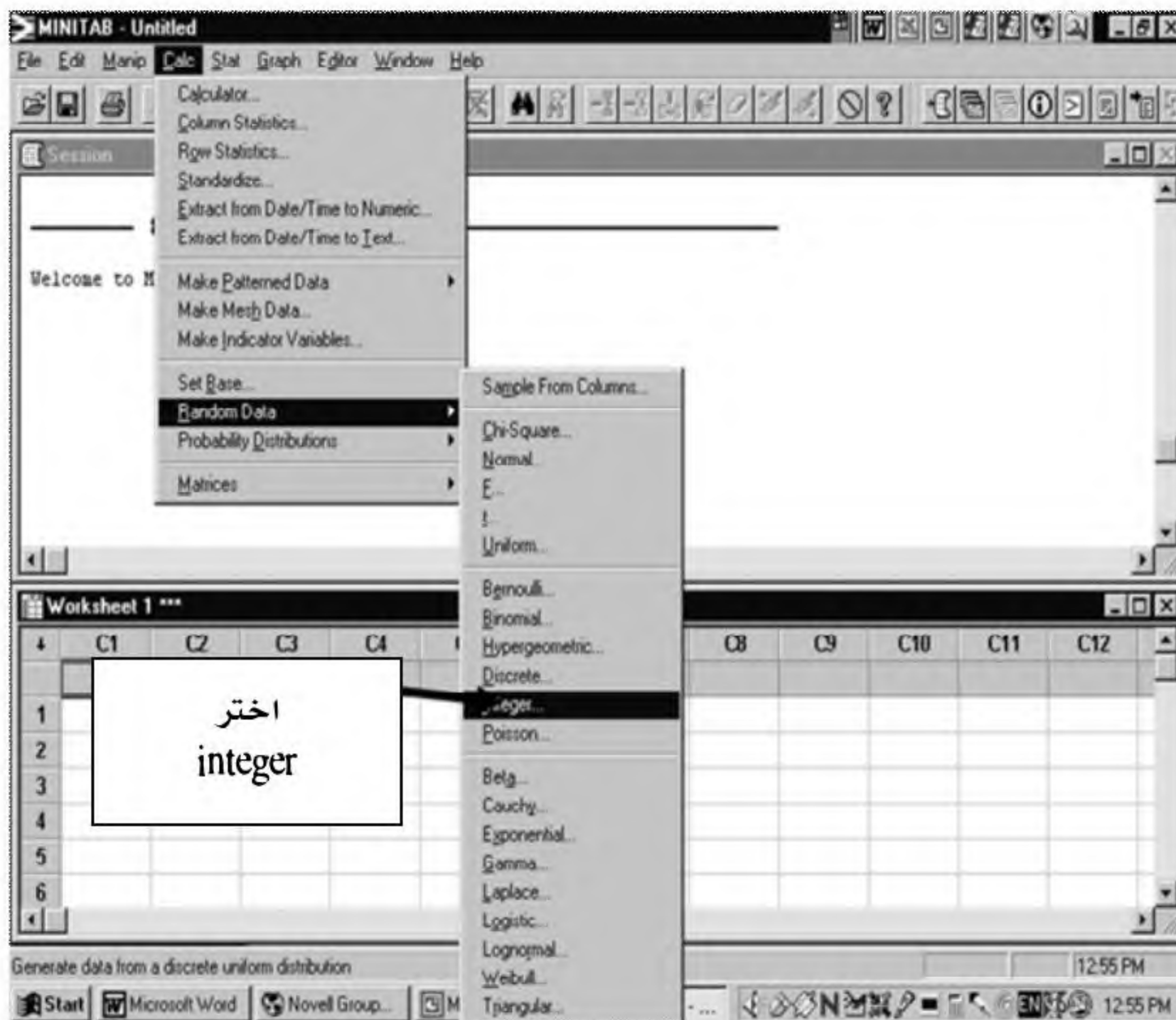
مثال: ترغب إحدى الشركات في سحب عشرة من العاملين فيها من بين عامل، لغرض جمع بعض البيانات، التي تتعلق بالخدامات المقدمة للعاملين في الشركة. جرى ترقيم العاملين من 1 إلى 300.

الحل: نقوم باستخدام الحقيبة الإحصائيةMINITAB، للحصول على عشرة أرقام بين 1 إلى 300، القيمة يجب أن تكون عدداً صحيحاً يقابل الرقم المعطى للعامل، بعد ترقيم العاملين في الشركة من 1 إلى 300، نقوم باتباع الخطوات الآتية لسحب العينة:

الخطوة الأولى:



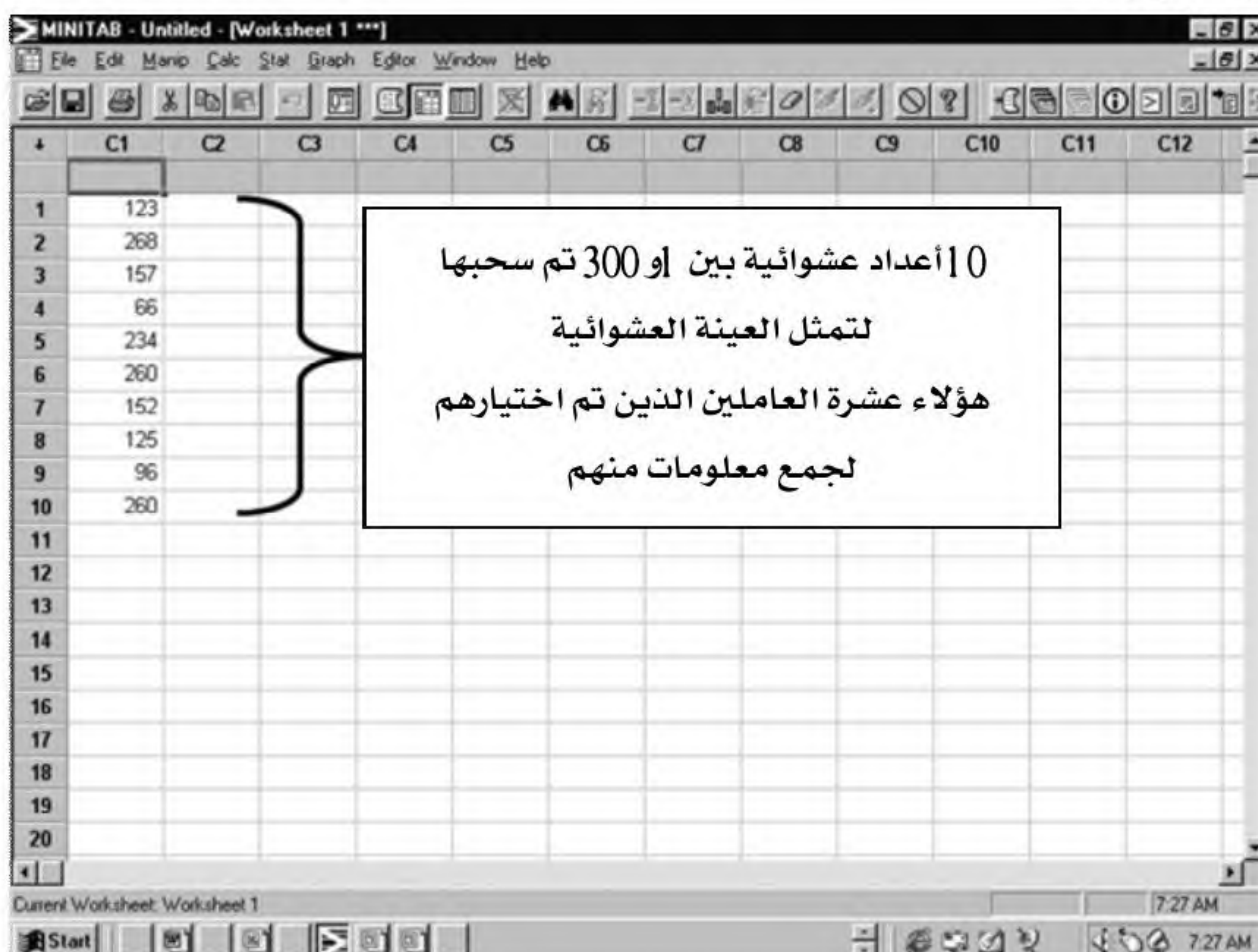
الخطوة الثانية:



الخطوة الثالثة:



الخطوة الرابعة:



الجزء الثاني

الجداول والمدرجات التكرارية والرسومات البيانية لوصف البيانات الإحصائية

أولاً: التوزيعات التكرارية (Frequency distribution)

مثال: يرغب أحد النوادي الرياضية برسم توزيع تكراري لعينة عشوائية حجمها 100 منتسب، سحبت من بين منتسبي النادي، وسئلوا عن عدد الأحذية التي يمتلكونها والمصنوعة من قبل شركة نايكي (Nike).

الحل: نقوم بخزن البيانات بأحد الأعمدة الموجودة في النافذة السفلى لحقيبةMINITAB أو نفتح الملف الذي يحتوي على هذه البيانات. في المثال الذي سنستخدمه، جرى خزن البيانات الخاصة بعدد الأحذية التي يمتلكونها والمصنعة من قبل شركة نايك (Nike). في العمود الخامس (C5). سنقوم باستخدام الحقيبة الإحصائية MINITAB لإيجاد التوزيع التكراري وفق الخطوات الآتية:

الخطوة الأولى:

MINITAB - Untitled - [SportsShoes.MTW ***]

File Edit Manip Calc Stat Graph Editor Window Help

↓	C1	C2	C3	C4	C5	C6	C7	C8
	Age	Annual Income	Pairs per year	Pairs at this time	Number of Nike	Number of Adidas	Brand Preference	Comf
3	18	8000	1	1	0	1	2	
4	48	33000	1	7	3	2	1	
5	37	27000	5	8	2	5	2	
6	56	40000	10					
7	22	11000	2					
8	57	56000	1					
9	24	38000	5					
10	63	9000	4					
11	34	24000	1					
12	21	12000	3					
13	29	15000	6					
14	35	23000	2	4	1	2	2	
15	31	19000					2	
16	46	52000					1	
17	27	31000					1	
18	49	62000					1	
19	19	8000					1	
20	22	13000					1	
21	24	11000					2	
22	74	15000					1	

ادخل البيانات الخاصة بعدد الأحذية المملوكة من قبل منتسبي النادي والمصنعة من قبل شركة نايكي أو افتح الملف الذي يحتوي على البيانات

الملف يحتوي على 100 مشاهدة في العمود الخامس الذي يحتوي على عدد الأحذية المملوكة من قبل منتسبي النادي والمصنوعة من قبل شركة نايكي

Current Worksheet: SportsShoes.MTW

7:44 AM

الخطوة الثانية:

MINITAB - Untitled - [SportsShoes.MTW ***]

File Edit Manip Calc Stat Graph Editor Window Help

↓	C1			C4	C5	C6	C7	C
	Age	Annual		Pairs at this time	Number of Nike	Number of Adidas	Brand Preference	Com
100	51			5	4	1	1	
101								
102								
103								
104								
105								
106								
107								
108								
109								
110								
111								
112								
113								
114								
115								
116								
117								
118								
119								

Basic Statistics
Regression
ANOVA
DOE
Control Charts
Quality Tools
Reliability/Survival
Multivariate
Time Series
Tables
Nonparametrics
EDA
Power and Sample Size

Cross Tabulation...
Tally...
Chi-Square Test...
Simple Correspondence Analysis...
Multiple Correspondence Analysis...

اختر Stat ومن ثم Tables ومن ثم Tally

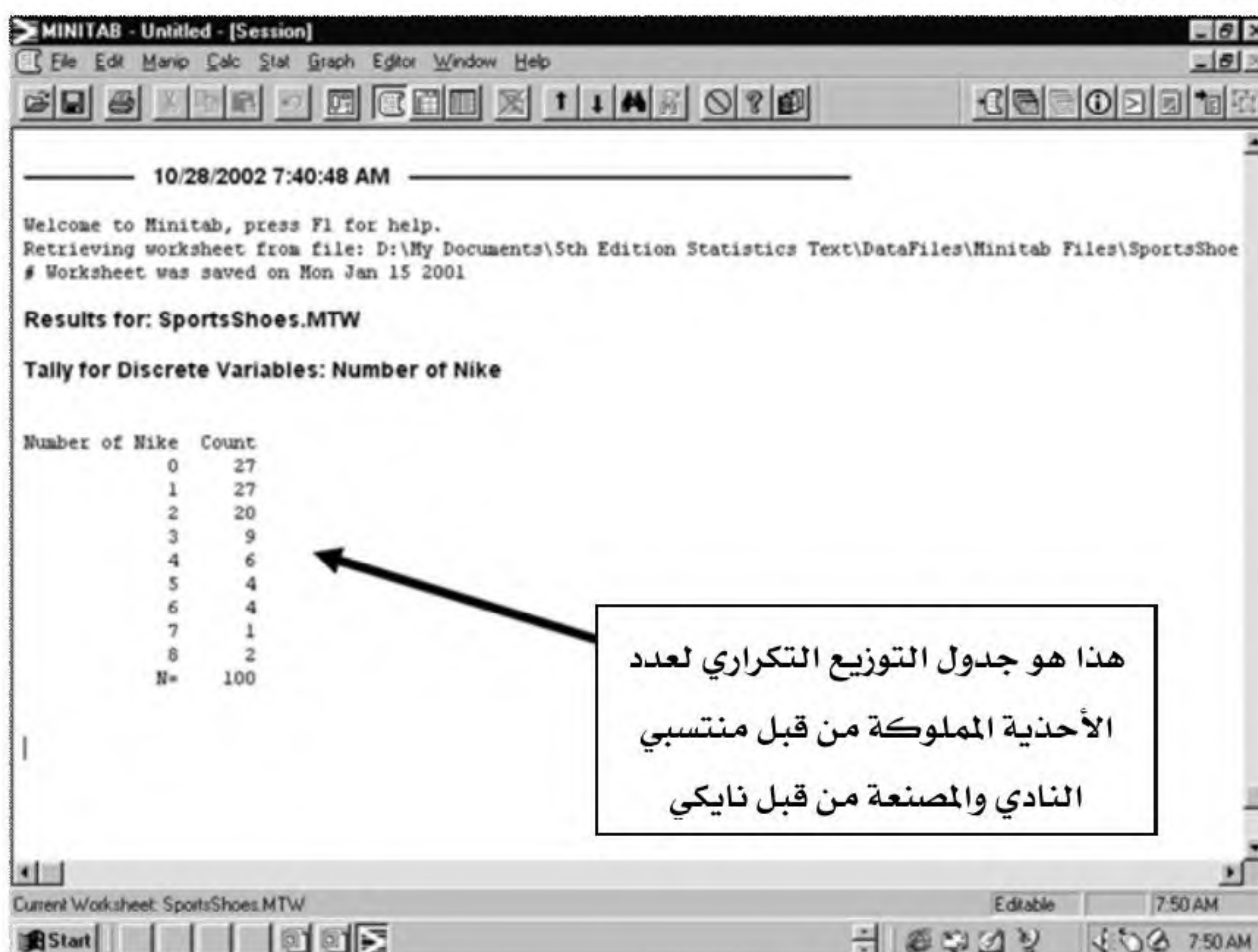
Display a one-way table of counts and percents for specified variables

7:47 AM

الخطوة الثالثة:



الخطوة الرابعة:



ثانياً: المدرجات التكرارية (Histogram)

مثال: يرغب أحد البنوك في تحليل رصيد بطاقة الائتمان لعملائه، قام بسحب عينة عشوائية من 300 عميل. قام بإدخال رقم الحساب في العمود الأول، وفي العمود الثاني الرصيد لبطاقة الائتمان الخاصة بهؤلاء العملاء، المطلوب بناء توزيع تكراري لكي يتسنى لنا تحليل هذه البيانات.

الحل: بعد إدخال البيانات الخاصة برصيد بطاقة الائتمان، نتبع الخطوات الآتية لبناء التوزيع التكراري:

الخطوة الأولى:

MINITAB - Untitled - [Capital.MTW ***]

File Edit Manip Calc Stat Graph Editor Window Help

	C1	C2	C3	C4	C5	
↓	Customer Number	Credit Card Account Balance	Gender 1 = M			Capital Cred
1	51130	10				
2	61413	10				
3	54713	10				
4	79247	3				
5	94899	7				
6	48035	4				
7	81473	7				
8	47715	1013	1			
9	77905	440	1			
10	87443	1246	2			
11	67666	1254				
12	51470	191				
13	79898	797				
14	98615	1127				
15	54278	646				
16	58800	1212	2			
17	73680	1036	1			
18	75732	477	1			
19	48393	591	1			
20	49325	886	1			

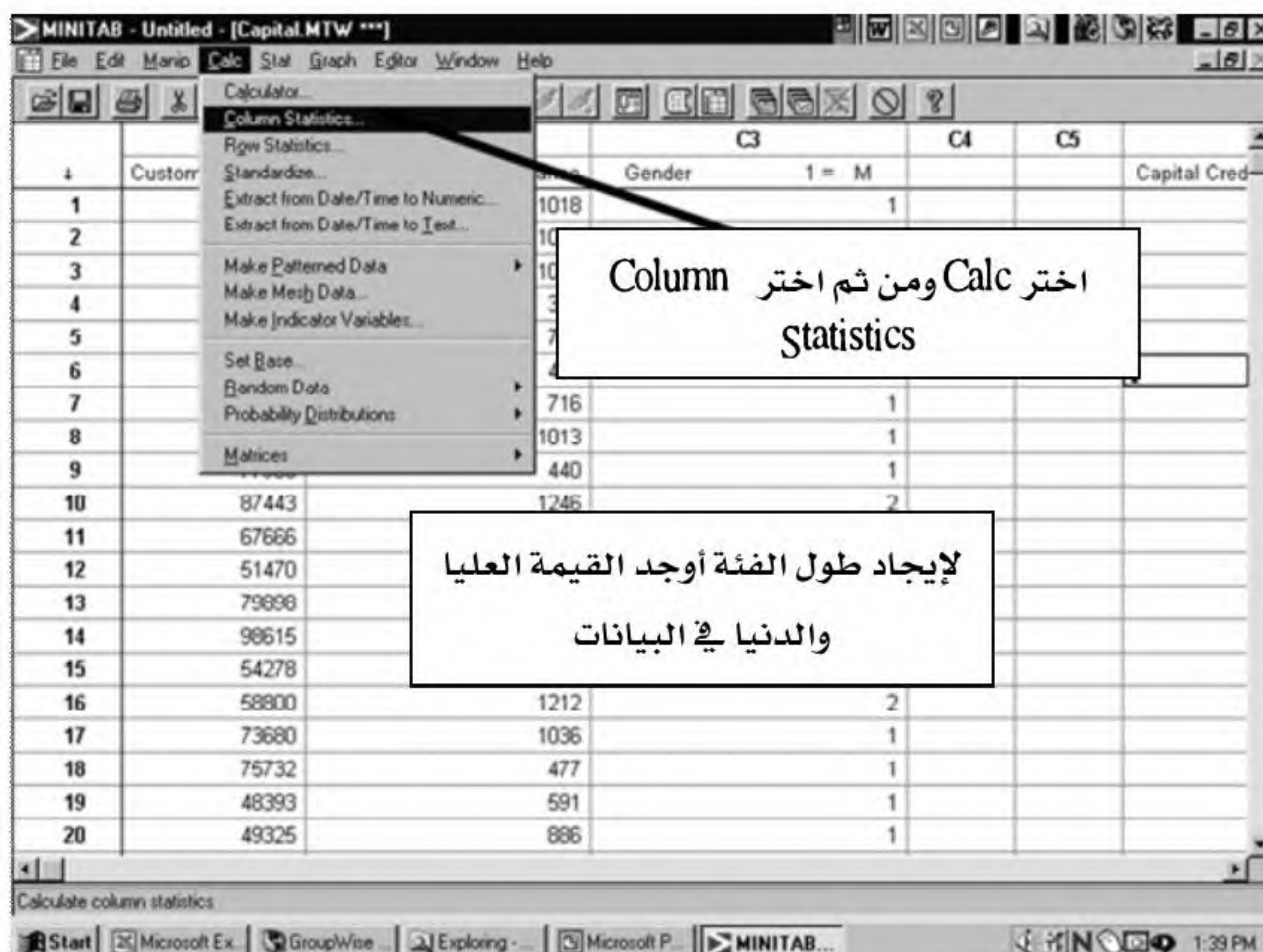
ادخل البيانات الخاصة برصيد بطاقة الائتمان لأحد البنوك أو افتح الملف الذي يحتوي على البيانات

الملف يحتوي على 300 مشاهدة العمود الثاني يحتوي رصيد بطاقة الائتمان

Current Worksheet: Capital.MTW 1:35 PM

Start | Microsoft Ex... | GroupWise ... | Exploring - ... | Microsoft P... | MINITAB... 1:35 PM

الخطوة الثانية:



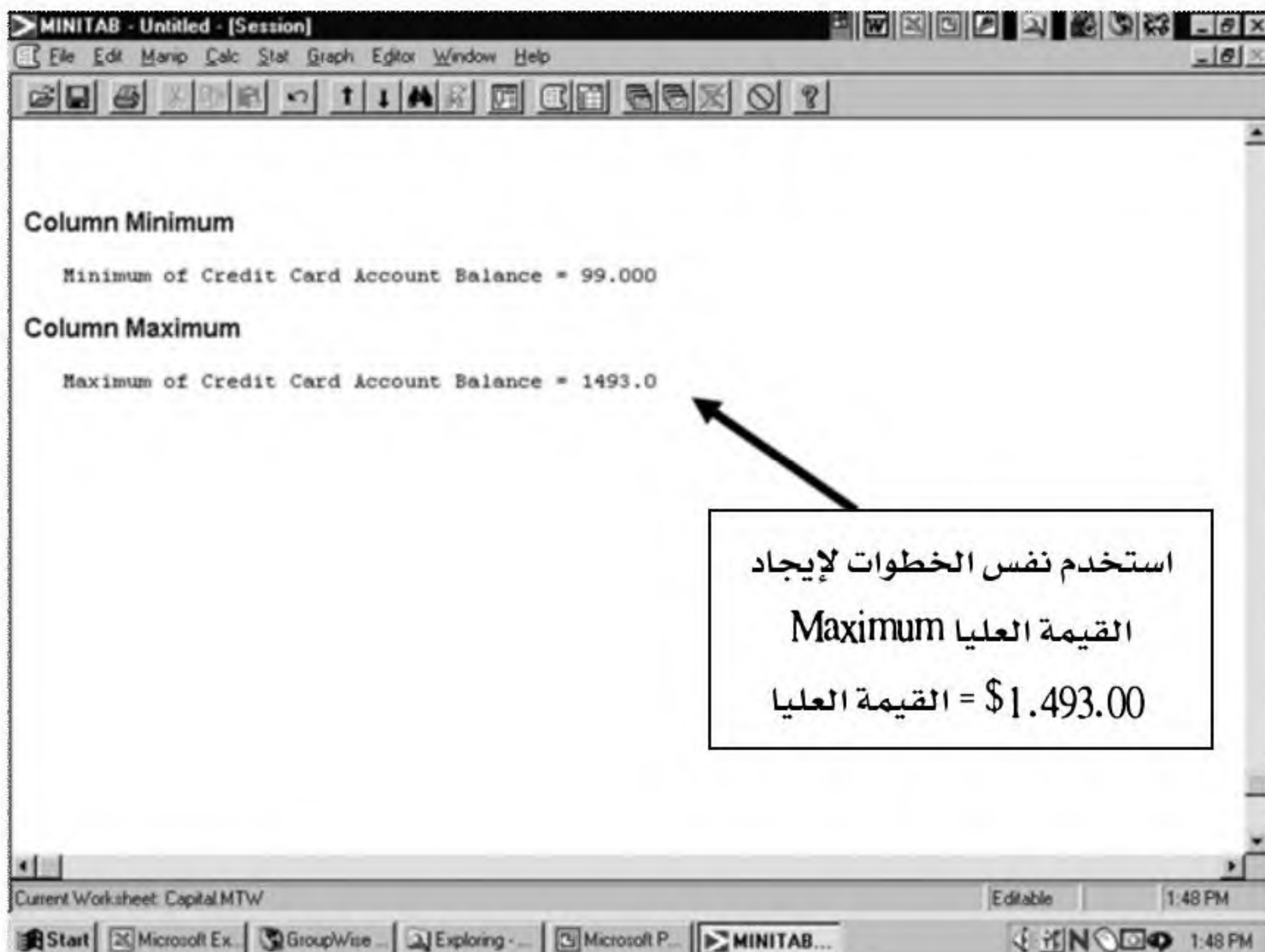
الخطوة الثالثة:



الخطوة الرابعة:



الخطوة الخامسة:



الخطوة السادسة:

MINITAB - Untitled - [Session]
File Edit Manip Calc Stat Graph Editor Window Help

Column Minimum

Minimum of Credit Card Account Balance = 99.000

Column Maximum

Maximum of Credit Card Account Balance = 1493.0

احسب طول الفئة

$$w = \frac{1493.00 - 99.00}{10} = 139.4$$

$150.00 قرب نحو الأعلى

الفئات

$90 إلى أقل من $240

$240 إلى أقل من $390

$390 إلى أقل من $540

etc□

Current Worksheet: Capital.MTW — Editable — 1:49 PM

Start | Microsoft Ex... | GroupWise ... | Exploring - ... | Microsoft P... | MINITAB... — 1:50 PM

الخطوة السابعة:

MINITAB - Untitled - [Capital.MTW ***]
File Edit Manip Calc Stat Graph Editor Window Help

↓	C6	C7-T	C8-T	C9	C10
	Capital Credit Card Customers			Cutpoints	
1				90	
2		File: Capital.mtw		240	
3				390	
4			Column C1: Customer Number	540	
5			Column C2: Credit Card Balance	690	
6			Column C3: Gender Code (1 = Male, 2 = Female)	840	
7				990	
8				1140	
9				1290	
10				1440	
11				1590	
12					
13					
14					
15					
16					
17					
18					
19					
20					

ادخل الحدود الدنيا للفئات

Current Worksheet: Capital.MTW — 11:48 AM

Start | Micros... | Micros... | MINI... | (Archiv... | Micros... — 11:48 AM

الخطوة الثامنة:



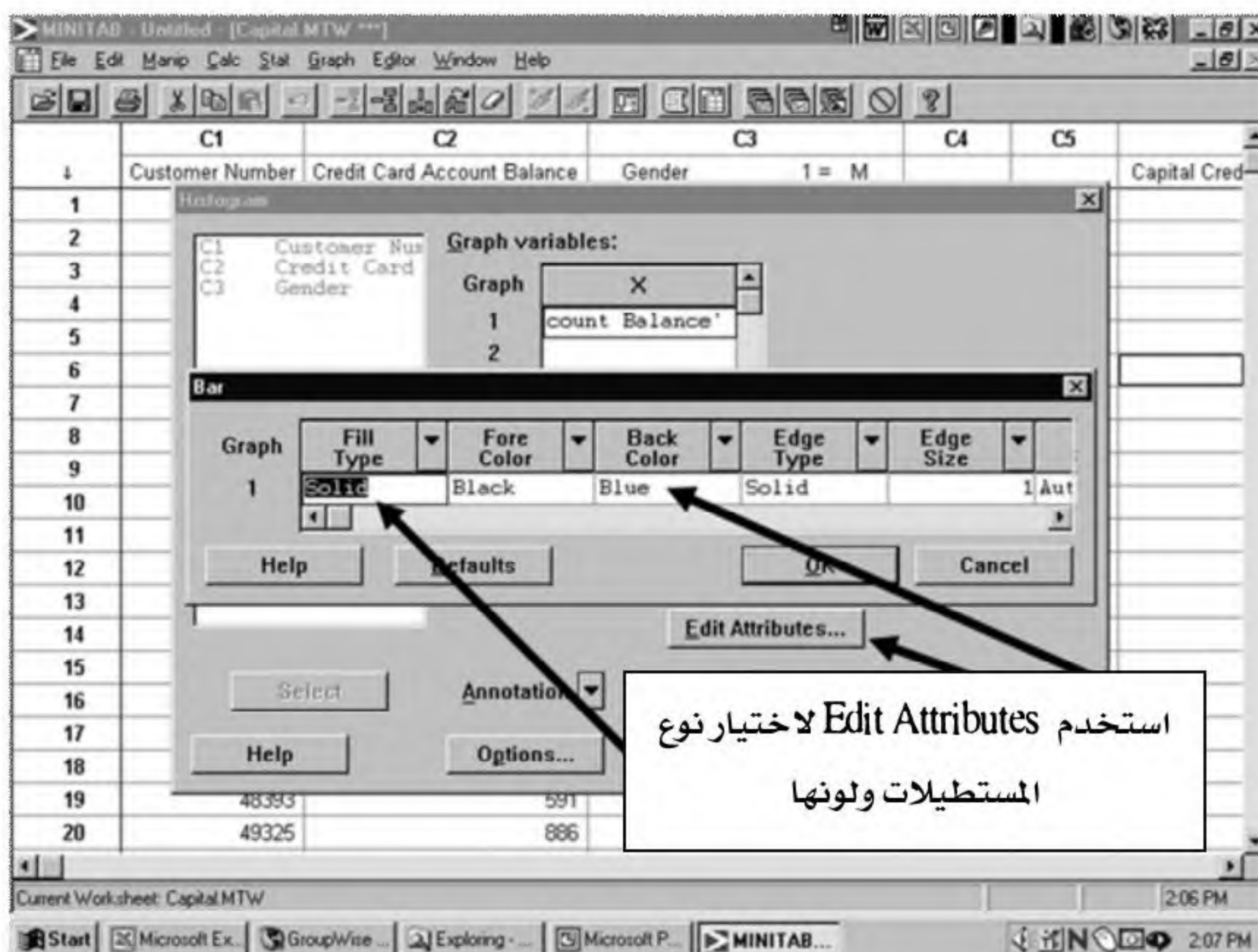
الخطوة التاسعة:



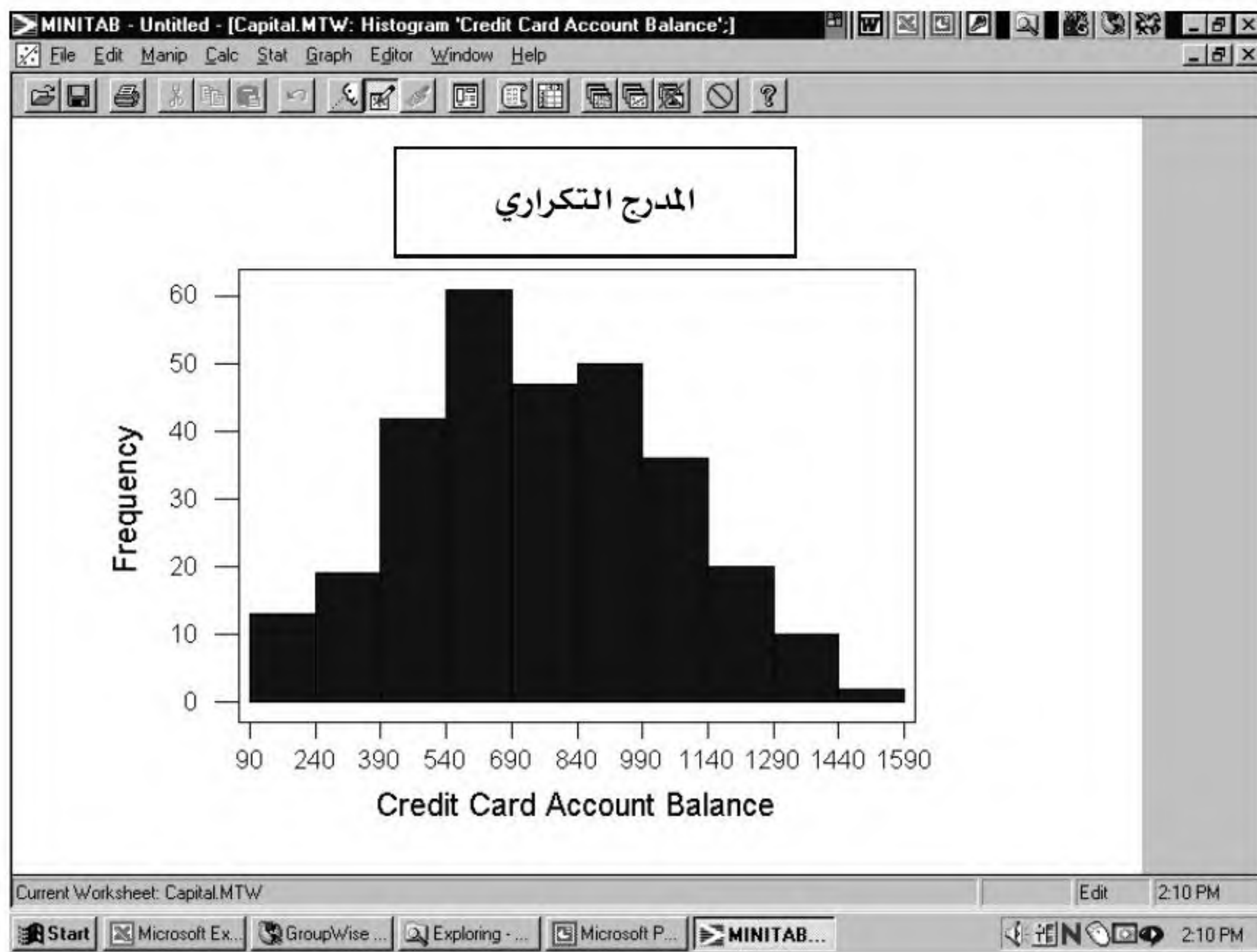
الخطوة العاشرة:



الخطوة الحادية عشرة:



الخطوة الثانية عشرة:



ثالثاً: المستطيلات البيانية (Bar charts)

مثال: رُفعت دعوى قضائية ضد إحدى الشركات، مفادها أن الشركة تدفع مرتبات للنساء الجدد اللاتي يتم تعيينهن في الشركة أقل من الرجال. قامت الشركة بتقديم كشف بمعدل المرتبات التي تم دفعها لكلا الجنسين ولسبع سنوات متتابعة. وترغب في رسم المستطيلات البيانية للمقارنة بين مرتبات العاملين فيها من الجنسين.

الحل: نقوم بإدخال البيانات بعمودين مختلفين، أو نفتح الملف الذي يحتوي على البيانات، ومن ثم نقوم برسم المستطيلات البيانية باتباع الخطوات الآتية:

الخطوة الأولى:

MINITAB - Untitled - [Bach.MTW ***]

File Edit Manip Calc Stat Graph Editor Window Help

↓	C1	C2	C3	C4	
	Year	Males: Average Starting Salarie	Females Average Starting Salari	Males: Percent with MBA	Females: Pe
1	1992	44456	41789	35	
2	1993	47286	46478	39	
3	1994	56234	53854	49	
4	1995	57890	58600	40	
5	1996	63467	59070	46	
6	1997	61090	55321	32	
7	1998	67543	64506	48	

افتح الملف الذي يحتوي على البيانات أو
ادخلها إن لم تكن محفوظة في ملف

Current Worksheet: Bach.MTW 8:19 PM

Start | GroupWise - Mailbox | Microsoft PowerPoint - | MINITAB - Untitle... 8:19 PM

الخطوة الثانية:

MINITAB - Untitled

File Edit Manip Calc Stat Graph Editor Window Help

Bach.MTW ***

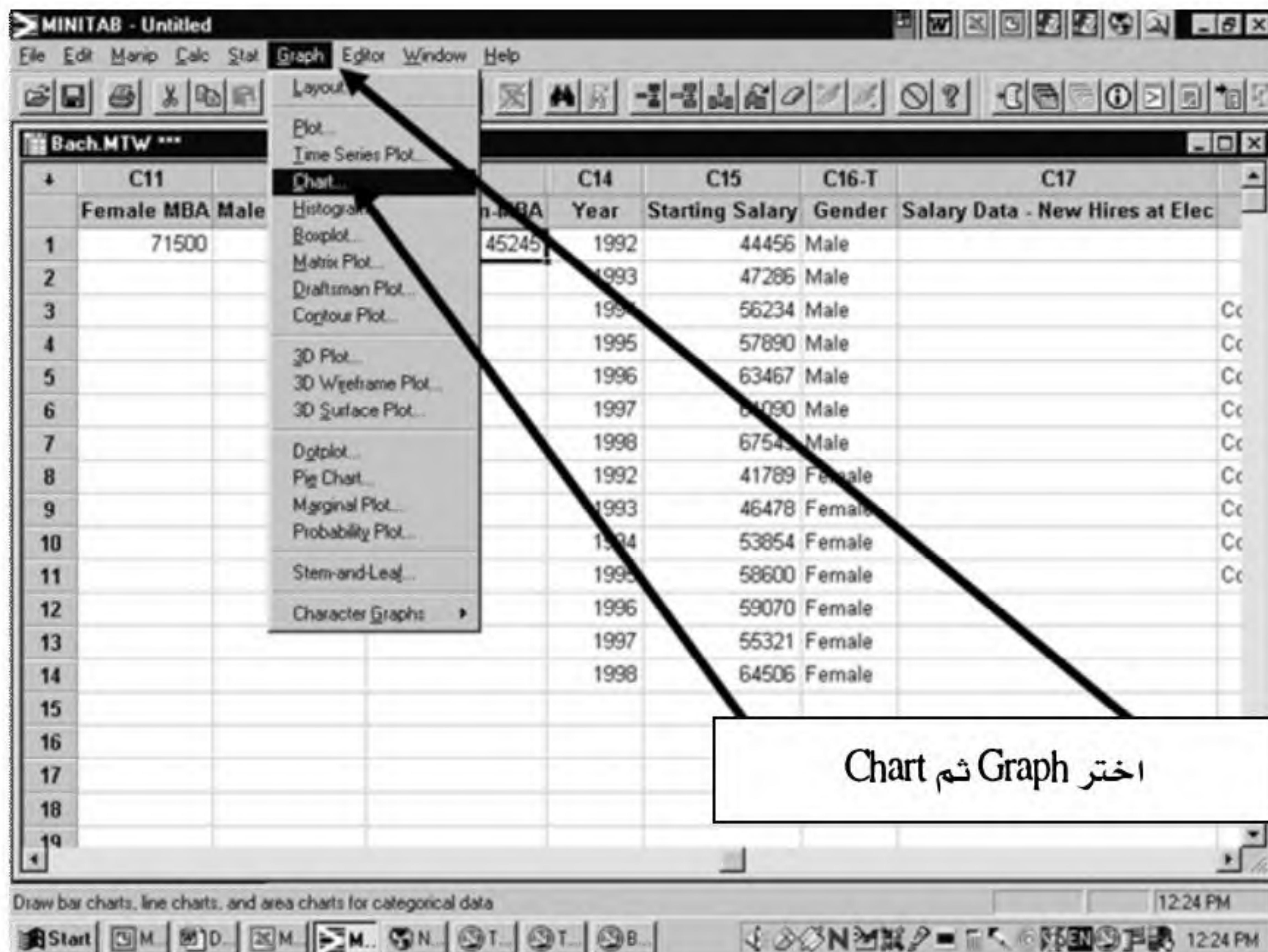
↓	C11	C12	C13	C14	C15	C16-T	C17
	Female MBA	Male non-MBA	Female non-MBA	Year	Starting Salary	Gender	Salary Data - New Hires at Elec
1	71500	43700	45245	1992	44456	Male	
2				1993	47286	Male	
3				1994	56234	Male	Cc
4				1995	57890	Male	Cc
5				1996	63467	Male	Cc
6					61090	Male	Cc
7					67543	Male	Cc
8					41789	Female	Cc
9					46478	Female	Cc
10					53854	Female	Cc
11					58600	Female	Cc
12					59070	Female	
13					55321	Female	
14				1998	64506	Female	

يجب أن تكون البيانات بهذا الشكل حتى
نستطيع أن نجد المستطيلات البيانية
باستخدام Minitab

Current Worksheet: Bach.MTW 11:38 AM

Start | Microsoft | Docume... | Microsoft | MINIT... 11:38 AM

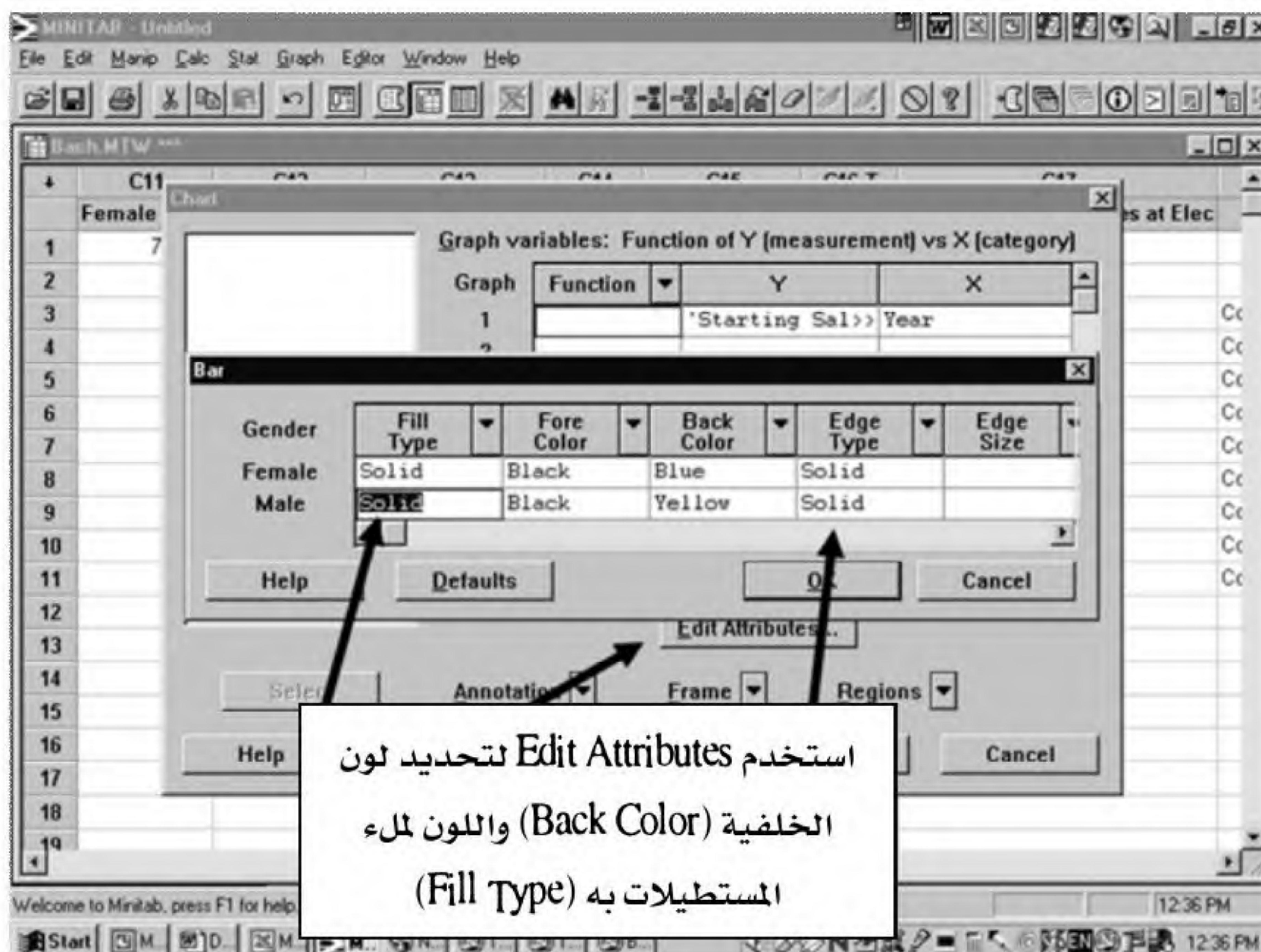
الخطوة الثالثة:



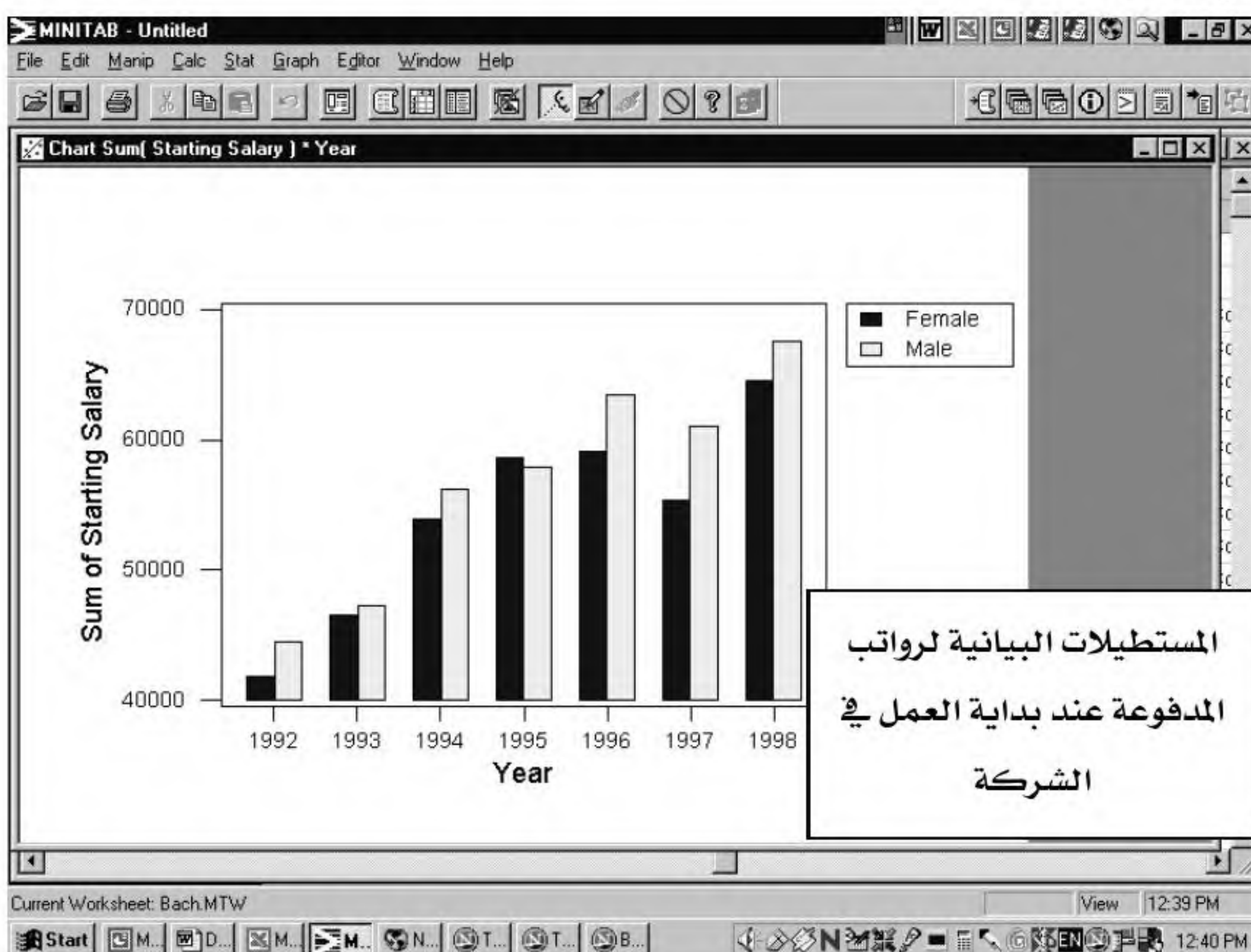
الخطوة الرابعة:



الخطوة الخامسة:



الخطوة السادسة:



رابعاً: الخطوط البيانية (Line charts)

مثال: ترغب إحدى الشركات في تحليل البيانات الخاصة بالمبيعات والأرباح لمدة زمنية محددة. قامت الشركة بإدخال البيانات في العمودين الثاني والثالث.

الحل: سنقوم بتحليل هذه البيانات باستخدام الخطوط البيانية وفق الخطوات الآتية:

الخطوة الأولى:

MINITAB - Untitled - [McGregor.MTW ***]

File Edit Manip Calc Stat Graph Editor Window Help

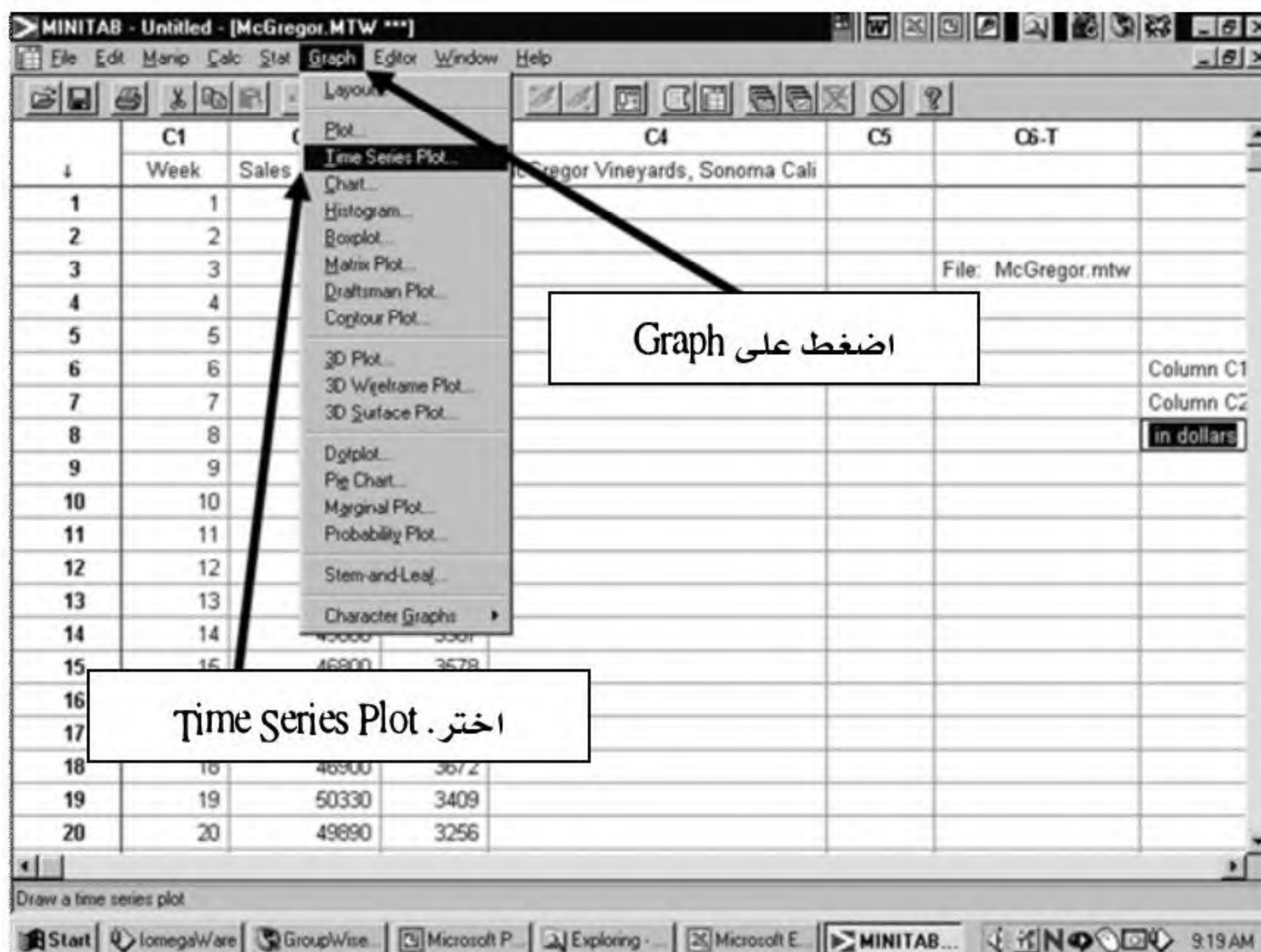
	C1	C2	C3	C4	C5	C6-T	
↓	Week	Sales (dollars)	Profit	McGregor Vineyards, Sonoma Cali			
1	1	37000	4120				
2	2	38900	4200				
3	3	41200	3960			File: McGregor.mtw	
4	4	39600	4010				
5	5	40500	3950				
6	6	42000	3875				Column C1
7	7	43500	3800				Column C2
8	8	41500	3910				in dollars
9	9	44100	3602				
10	10	42800	3500				
11	11	43700	3680				
12	12	47900	3476				
13	13	46790	3680				
14	14	49000	3367				
15	15	46800	3578				
16	16	48680	3456				
17	17	50100	3345				
18	18	46900	3672				
19	19	50330	3409				
20	20	49890	3256				

افتح الملف الذي يحتوي على البيانات أو ادخلها، البيانات التي سنستخدمها تحتوي على بيانات لعشرين أسبوعًا

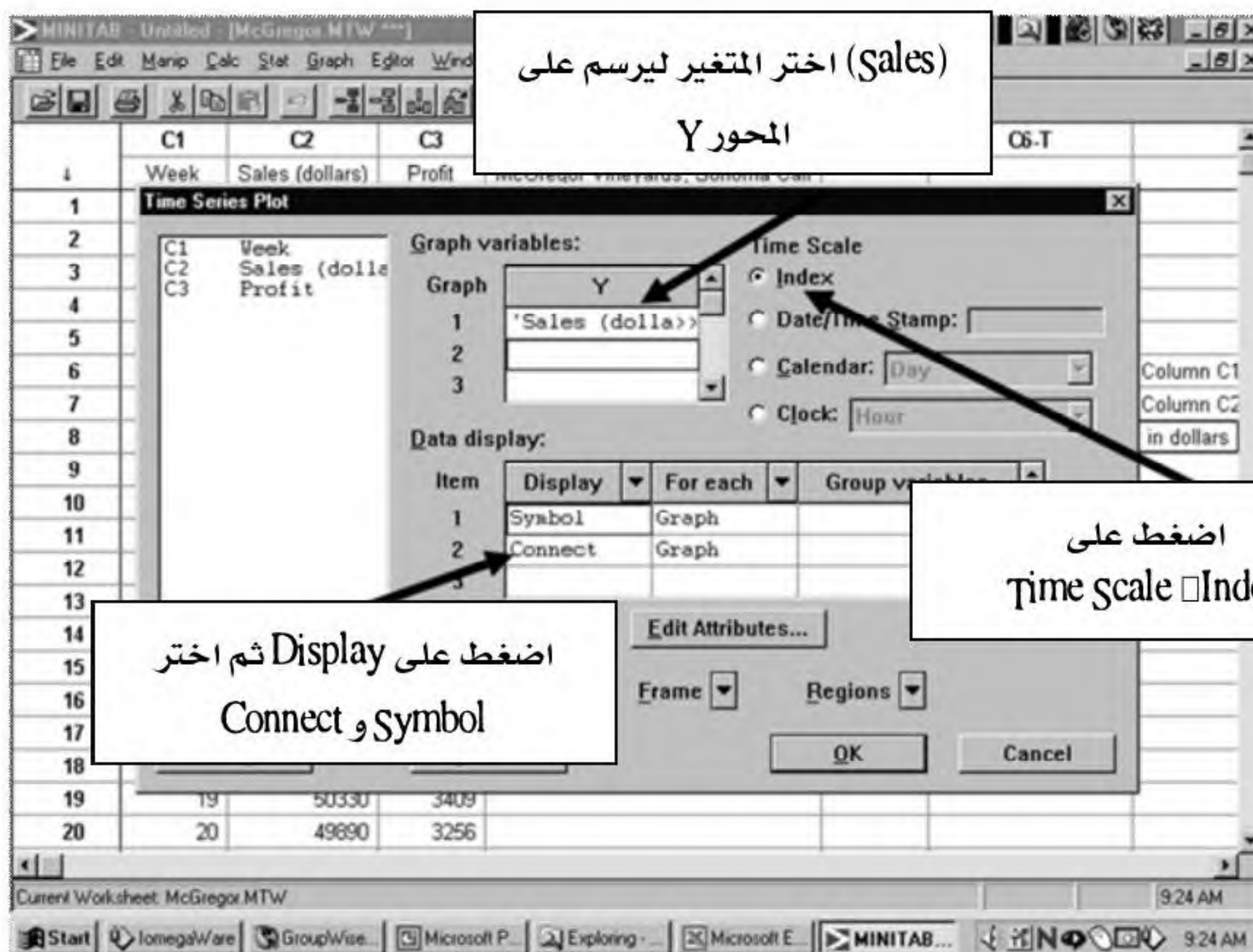
Current Worksheet: McGregor.MTW 9:18 AM

Start | IomegaWare | GroupWise... | Microsoft P... | Exploring - ... | Microsoft E... | MINITAB... 9:18 AM

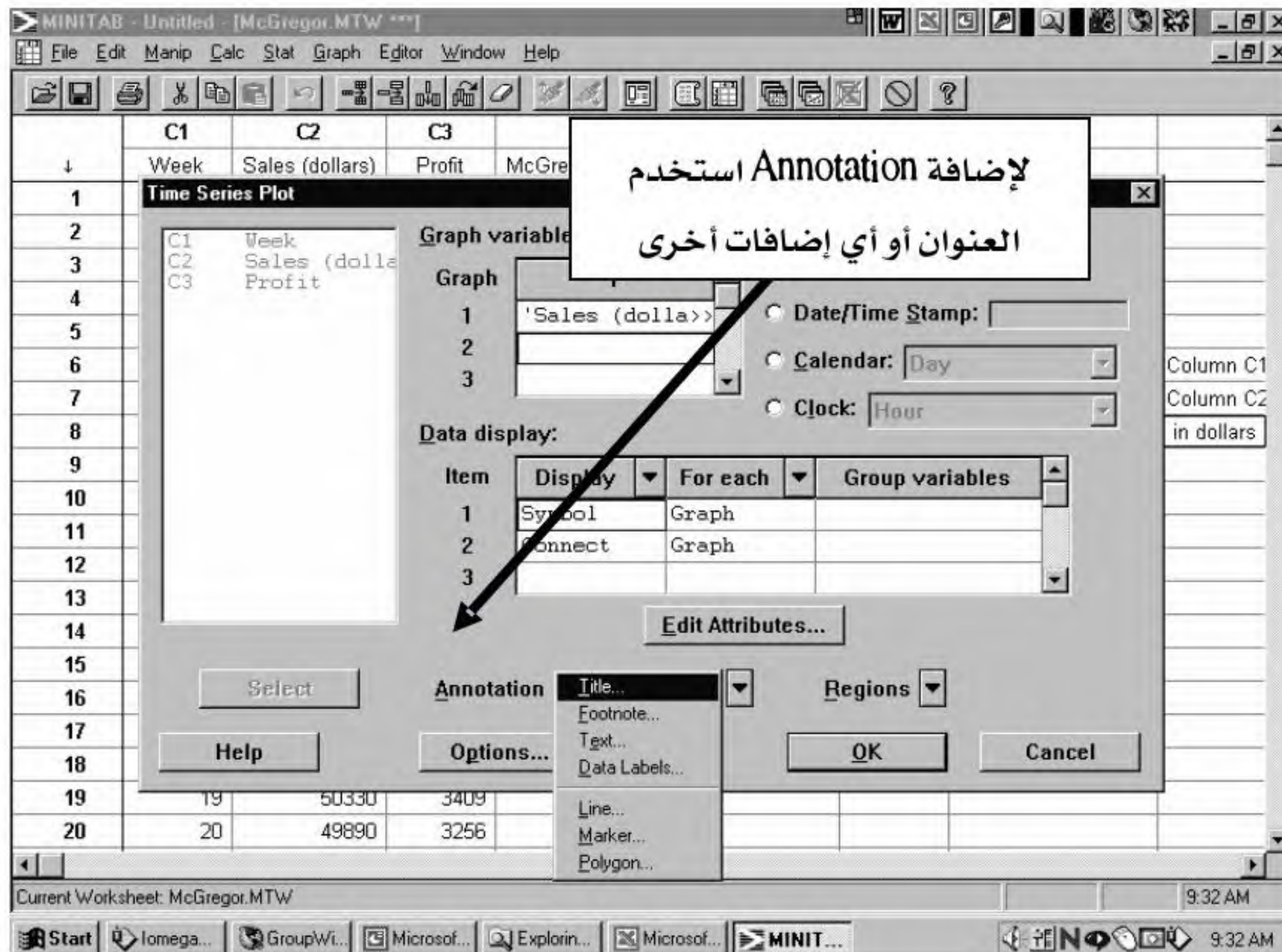
الخطوة الثانية:



الخطوة الثالثة:



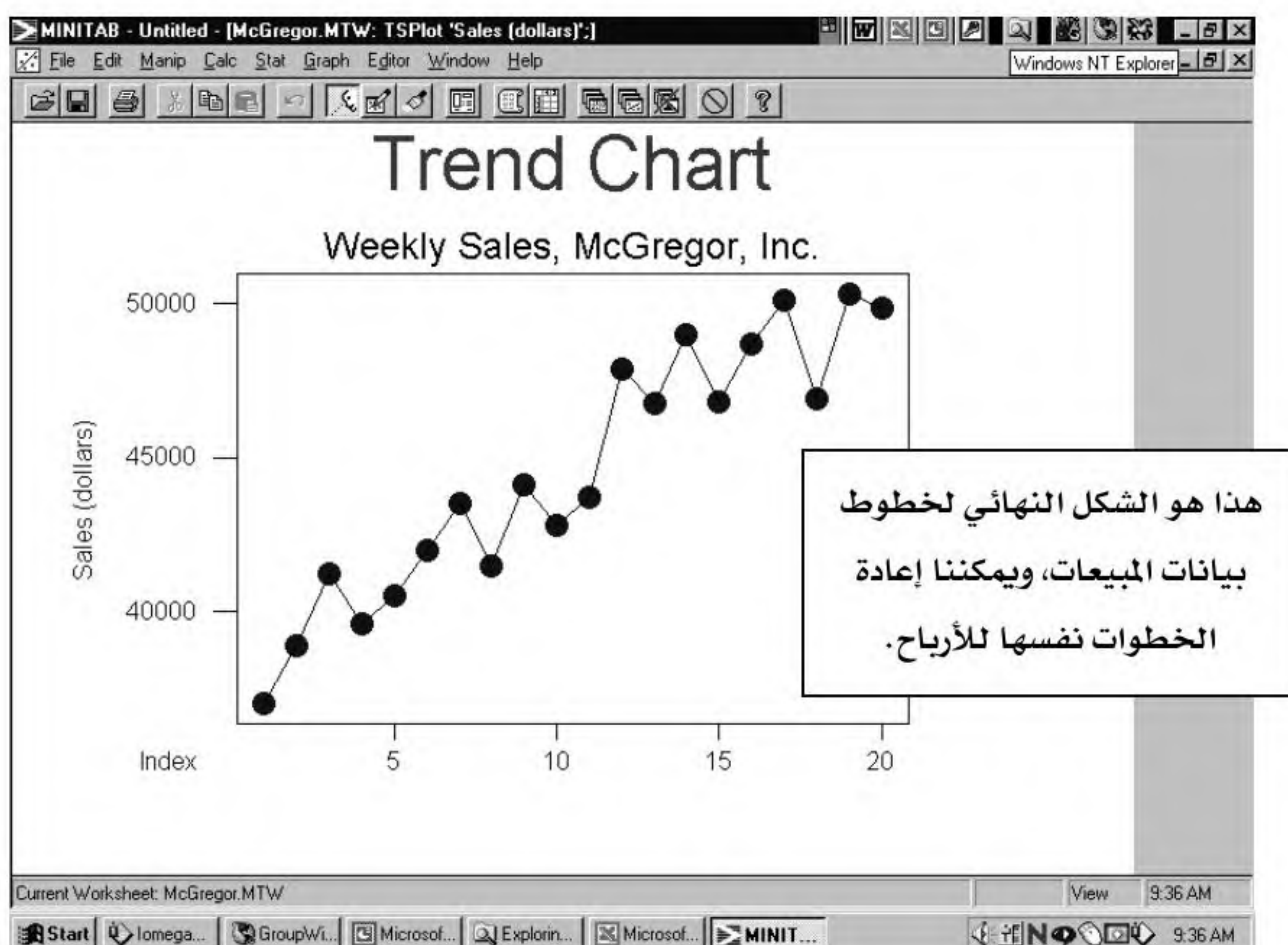
الخطوة الرابعة:



الخطوة الخامسة:



الخطوة السادسة:



خامساً: شكل الانتشار (scatter diagrams)

مثال: ترغب إحدى المحلات التجارية في دراسة العلاقة بين سعر الحاسب وسرعته، قامت بإدخال البيانات الخاصة بسعر الحاسب وسرعته في العمودين الثالث والخامس، بالإضافة إلى بيانات أخرى.

الحل: سنقوم بدراسة العلاقة بين سعر الحاسب وسرعته باستخدام شكل الانتشار بين المتغيرين السعر والسرعة وفق الخطوات الآتية:

الخطوة الأولى:

MINITAB - Untitled - [Computers.MTW ***]

File Edit Manip Calc Stat Graph Editor Window Help

	C1-T	C2-T	C3	C4-T	C5	C6	C7
↓	Brand	Model	Price	Includes Monitor	Processor Mz	RAM	Hard Drive C
1	All Computer Technology	EcoPower	998	Yes	333	32	
2	American Multisystems	InfoGold	995	Yes	300	32	
3	Apple Computer	iMac	1299	Yes	233	32	
4	AST	Bravo Elc	799	No	300	32	
5	Compaq	Prosignia	1239	Yes	300	32	
6	CompUSA	PC Ceneron	930	Yes	300	32	
7	Gateway	E1200-333				32	
8	Hewlett-Packard	Brio 7100				64	
9	Hewlett-Packard	Pavillion				64	
10	IBM	Aptiva E2N				32	
11	IBM	PC-300GL				32	
12	Magitronic Technology	Pardigm 2000				64	
13	Packard Bell	Multimedia				64	
14	Quantex Microsystems	QP6/266				32	
15	Tiger Direct	K-Series				64	
16	Tochiba American Information	Equium 7100M				32	
17	All Computer Technology	Titum 6450	2988	Yes	400	64	
18	AMAX Engineering	Microplex	3099	Yes	450	128	
19	American Multisystems	InfoGold PII 400	1895	Yes	400	64	10.2
20	Apple Computer	Power Mac	2999	No	333	128	

Current Worksheet: Computers.MTW 11:02 AM

افتح الملف الذي يحتوي على البيانات أو ادخلها. البيانات التي سنستخدمها تحتوي على بيانات 36 حاسبًا

الخطوة الثانية:

MINITAB - Untitled - [Computers.MTW ***]

اضغط على Graph ومن ثم اختر Character Graphs وScatter plot

Graph: End Layout, Plot..., Time Series Plot..., Chart..., Histogram..., Boxplot..., Matrix Plot..., Draftsman Plot..., Contour Plot..., 3D Plot..., 3D Wireframe Plot..., 3D Surface Plot..., Dotplot..., Pie Chart..., Marginal Plot..., Probability Plot..., Stem-and-Leaf..., Character Graphs

Character Graphs: Set Options..., Histogram..., Boxplot..., Dotplot..., Stem-and-Leaf..., Scatter Plot..., Multiple Scatter Plot..., Time Series Plot..., Grid..., Contour..., Pseudo 3D Plot...

Draw a character-style scatter plot that displays in the Session window 12:57 PM

الخطوة الثالثة:



الجزء الثالث

1- مقاييس النزعة المركزية: الوسط والوسيط.

2- مقاييس الموضع: الربيعيات والمئينيات.

3- مقاييس التشتت: المدى والانحراف المعياري.

أولاً: مقاييس النزعة المركزية (الوسط والوسيط).

مثال: ترغب إحدى المؤسسات في إيجاد الوسط الحسابي والوسيط لمتغيرين هما Gross vim وGross Poe، هناك 200 مشاهدة لكل متغير، قامت الشركة بإدخال البيانات في العمودين الثاني والخامس.

الحل: هنالك طريقتان يمكن استخدامهما لاستخراج قيمتي الوسط والوسيط للمتغيرين، الأولى سبق وأن وضحنا كيفية استخدامها عندما تكلمنا عن استخراج القيمة الصغرى والكبرى لمجموعة من البيانات، في هذه الطريقة

يمكننا أن نختار الوسط (Mean) أو الوسيط (Median) بدلاً من القيم الصغرى أو الكبرى في مثالنا السابق. سوف نقوم باستخراج قيمة الوسط الحسابي والوسيط باستخدام طريقة ثانية وفق الخطوات الآتية:

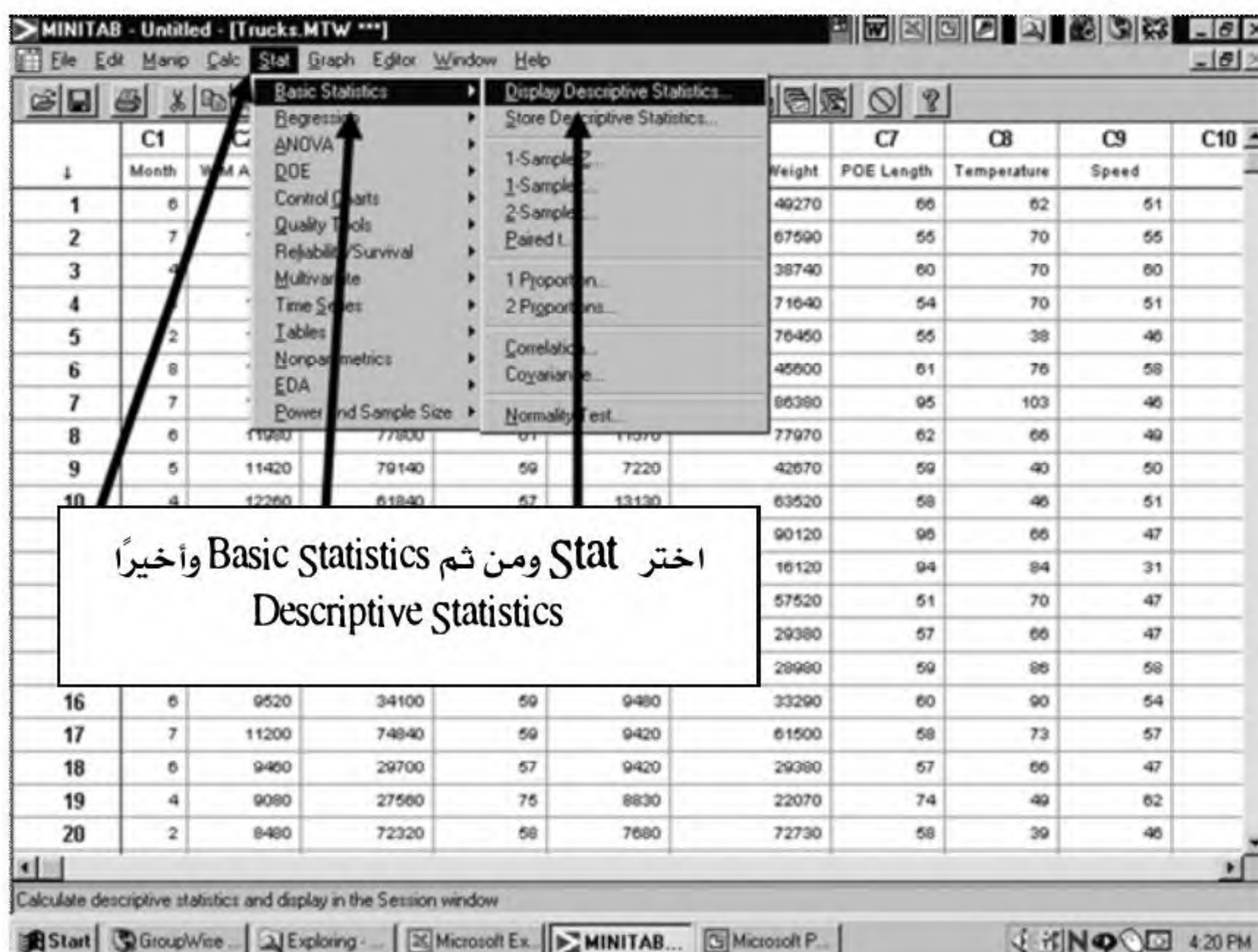
الخطوة الأولى:

MINITAB - Untitled - [Trucks.MTW ***]

File Edit Manip Calc Stat Graph Editor Window Help

↓	C1 Month	C2 WIM A Wgt.	C3 WIM Gross Wgt.	C4 WIM Length	C5 POE A Wgt.	C6 POE Gross Weight	C7 POE Length	C8 Temperature	C9 Speed	C10
1	6	9560	50100	64	8870	49270	66	62	51	
2	7	10180	62880	55	10780	67590	55	70	55	
3	4	9060	40560	59	10350	38740	60	70	60	
4	8	11460	72100	56	10940	[illegible]	[illegible]	[illegible]	[illegible]	
5	2	11420	74040	53	11320	[illegible]	[illegible]	[illegible]	[illegible]	
6	8	10480	42320	60	9680	[illegible]	[illegible]	[illegible]	[illegible]	
7	7	10700	82660	95	10930	[illegible]	[illegible]	[illegible]	[illegible]	
8	6	11980	77800	61	11570	[illegible]	[illegible]	[illegible]	[illegible]	
9	5	11420	79140	59	7220	[illegible]	[illegible]	[illegible]	[illegible]	
10	4	12260	61840	57	13130	[illegible]	[illegible]	[illegible]	[illegible]	
11	6	10460	95300	94	9290	[illegible]	[illegible]	[illegible]	[illegible]	
12	8	11700	108040	86	10346	16120	94	84	31	
13	4	8960	74340	51	7960	57520	51	70	47	
14	6	9460	29700	57	9420	[illegible]	[illegible]	[illegible]	[illegible]	
15	6	9320	32160	59	9210	[illegible]	[illegible]	[illegible]	[illegible]	
16	6	9520	34100	59	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	
17	7	11200	74840	[illegible]	9420	[illegible]	[illegible]	[illegible]	[illegible]	
18	6	9460	[illegible]	57	9420	[illegible]	[illegible]	[illegible]	[illegible]	
19	4	[illegible]	27560	75	8830	[illegible]	[illegible]	[illegible]	[illegible]	
20	2	8480	72320	58	7680	72730	58	39	46	

الخطوة الثانية:



الخطوة الثالثة:



الخطوة الرابعة:



ثانياً: مقاييس الموضع (الربيعيات والمئينيات)

مثال: ترغب المؤسسة في إيجاد الربيعيات والمئينيات للمتغيرين Vim Gross و Poe Gross للبيانات نفسها.

الحل: إذا كنا نرغب في إيجاد الربيع الأول والثالث، يمكننا أن نتبع الخطوات نفسها التي أوجدنا بها الوسط والوسيط ومن الخطوة الأخيرة نجد الربيع الأول والثالث كما هو موضح أدناه:

لإيجاد قيم **المئينيات** نتبع الخطوات الآتية:

الخطوة الأولى:

MINITAB - Untitled - [Trucks.MTW ***]

File Edit Manip Calc Stat Graph Editor Window Help

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
↓	Month	WIM A Wgt.	WIM Gross Wgt.	WIM Length	POE A Wgt.	POE Gross Weight	POE Length	Temperature	Speed	
1	6	9560	50100	64	8870	49270	66	62	51	
2	7	10180	62880	55	10780	67590	55	70	55	
3	4	9060	40560	59	10350	38740	60	70	60	
4	8	11460	72100	56	10940	71640	54	70	51	
5	2	11420	74040	53	11320					
6	8	10480	42320	60	9680					
7	7	10700	82660	95	10930					
8	6	11980	77800	61	11570					
9	5	11420	79140	59	7220					
10	4	12260	61840	57	13130					
11	6	10460	95300	94	9290	90120	96	66	47	
12	8	11700	108040	86	10346	16120	94	84	31	
13	4	8960	74340	51	7960					
14	6	9460	29700	57	9420					
15	6	9320	32160	59	9210					
16	6	9520	34100	59						
17	7	11200	74840		9420					
18	6	9460		57	9420					
19	4		27560	75	8830	22070	74	49	62	
20	2	8480	72320	58	7680	72730	58	39	46	

Current Worksheet: Trucks.MTW 3:42 PM

Start | GroupWise ... | Exploring - ... | Microsoft Ex... | MINITAB... | Microsoft P... 3:43 PM

ادخل البيانات التي ترغب بحساب الوسط الحسابي والوسيط لها

سوف نستخدم الملف الذي فيه 200 مشاهدة

الخطوة الثانية:

MINITAB - Untitled - [Trucks.MTW ***]

File Edit Manip Calc Stat Graph Editor Window Help

Basic Statistics ▸
Regression ▸
ANOVA ▸
DOE ▸
Control Charts ▸
Quality Tools ▸
Reliability/Survival ▸
Multivariate ▸
Time Series ▸
Tables ▸
Nonparametrics ▸
EDA ▸
Power and Sample Size ▸

Distribution ID Plot...
Distribution Overview Plot...
Parametric Dist Analysis-Right Cens...
Nonparametric Dist Analysis-Right Cens...
Parametric Dist Analysis-Arbitrary Cens...
Nonparametric Dist Analysis-Arbitrary Cens...
Accelerated Life Testing...
Regression with Life Data...
Probit Analysis...

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10
↓	Month			Length	POE A Wgt.	POE Gross Weight	POE Length	Temperature	Speed	
1	6			64	8870	49270	66	62	51	
2							55	70	55	
3	4						60	70	60	
4	8						54	70	51	
5	2						55	38	46	
6	8						61	76	58	
7	7						95	103	46	
8	6		77800				62	66	49	
9	5	11420	79140				59	40	50	
10	4	12260	61840				58	46	51	
							96	66	47	
							94	84	31	
							51	70	47	
							57	66	47	
							59	86	58	
16	6	9520	34100	59	9480	33290	60	90	54	
17	7	11200	74840	59	9420	61500	58	73	57	
18	6	9460	29700	57	9420	29380	57	66	47	
19	4	9080	27560	75	8830	22070	74	49	62	
20	2	8480	72320	58	7680	72730	58	39	46	

Perform a distribution analysis on uncensored/right censored data that follow a parametric distribution

Start | GroupWise - Ma... | Exploring - Busi... | Microsoft Excel ... | Microsoft Power... | MINITAB - U... 8:28 AM

اختر Stat ومن ثم Reliability/Survival ثم
Parametric Dist Analysis Right Cens

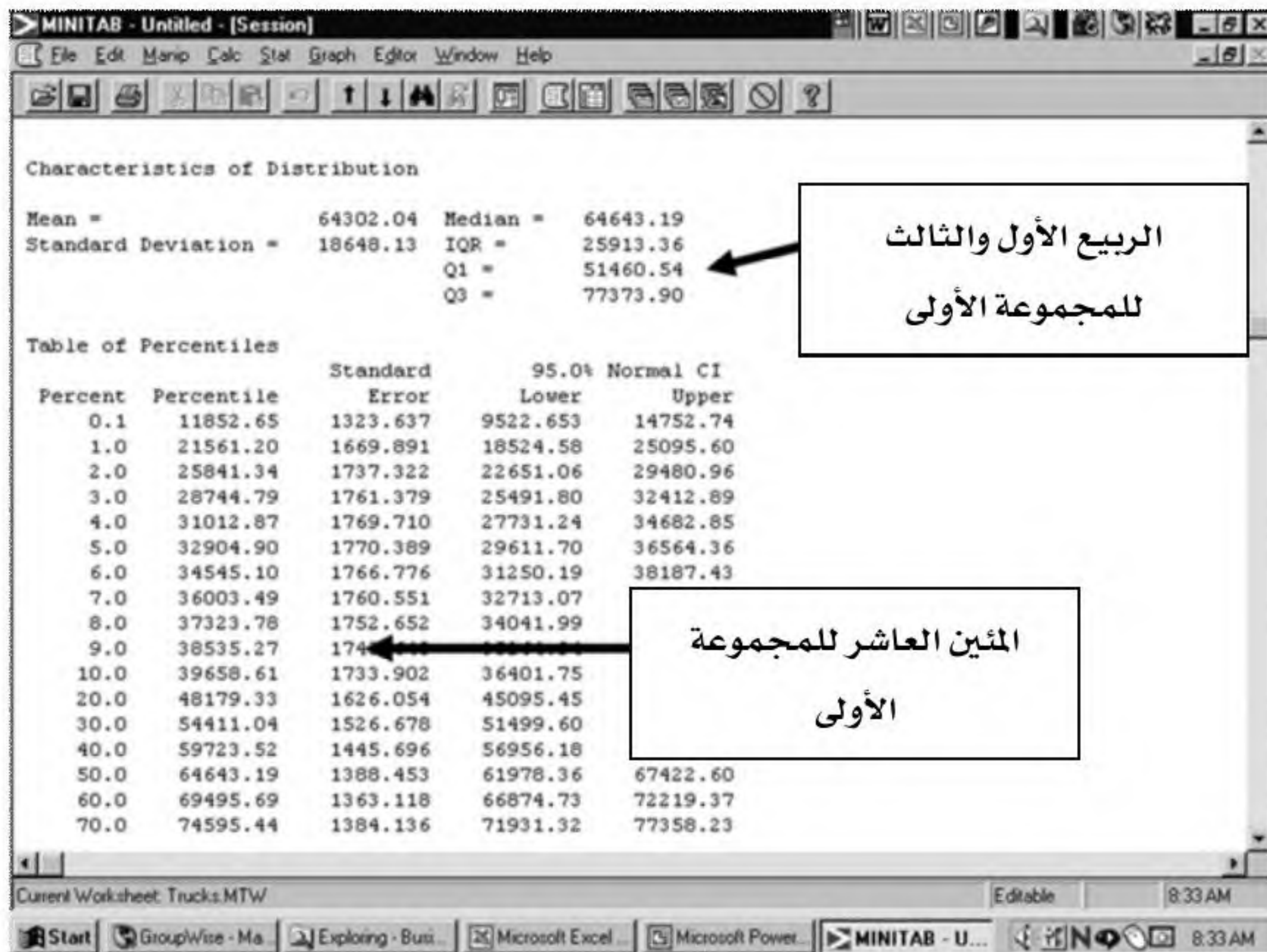
الخطوة الثالثة:



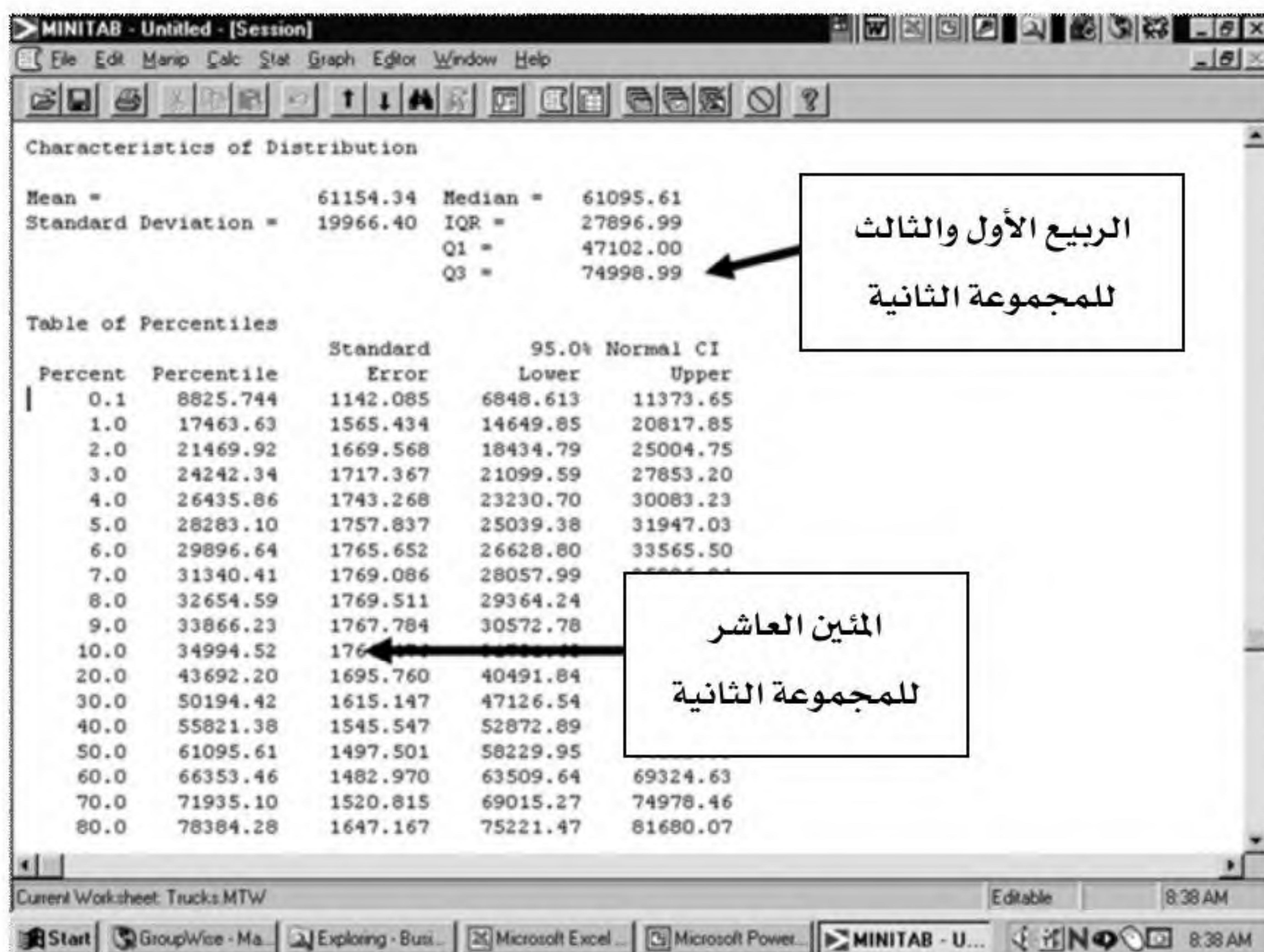
الخطوة الرابعة:



الخطوة الخامسة:



الخطوة السادسة:



ثالثاً: مقاييس التشتت (المدى والانحراف المعياري)

مثال: ترغب المؤسسة في إيجاد الانحراف المعياري والمدى للمتغيرين VimGross و Poe Gross للبيانات نفسها.

الحل: نتبع الخطوات نفسها التي اتبعناها لإيجاد الوسط والوسيط ولكن في الخطوة الأخيرة نجد قيمة الانحراف المعياري والمدى كما هو موضح أدناه.

الخطوة الرابعة:



جداول الأعداد العشوائية3 .

Random Number Table

	1	2	3	4	5	6	7	8	9	10	11	12	13
1	8571	7683	5118	7669	6126	3663	3059	7807	9219	4383	9021	7013	3348
2	4077	0864	5055	8631	0386	9792	1690	4874	3084	0228	8539	9375	8635
3	4753	1992	8182	2658	2914	4573	1499	7772	7759	9405	9502	7862	3716
4	1267	1063	4415	8496	6779	3408	6931	7946	4655	7715	9083	1714	6890
5	9497	0105	5626	0529	0602	7679	1726	1390	8186	7883	6180	5351	3961
6	6823	7365	6140	0357	7069	7470	0315	5422	8232	5943	1131	1577	3008
7	6532	4048	3044	8035	1045	2847	4715	2869	8349	6649	7059	4005	1563
8	8184	8336	5684	5846	7056	5605	0939	9380	5348	4997	9808	7833	4016
9	7299	4926	2348	5425	2722	7092	5893	1647	5373	0384	9803	1563	3070
10	2195	1561	5374	9679	7807	4437	7569	7307	2576	8175	9033	5632	6813
11	9526	2406	0715	5069	5908	6006	3449	7158	8950	6910	4166	1899	0252
12	7413	4884	0667	1663	7571	2142	5650	9061	7687	3849	2207	6754	1252
13	6131	8601	3370	8015	8299	2273	5475	6327	8295	8321	3075	9462	0674
13	0933	3524	9670	9120	9175	8937	1002	7659	6069	6695	4177	6170	7009
15	5265	9099	0391	6929	5171	3603	0309	0011	2902	6086	1054	3266	2278
16	4436	2960	8108	2014	5500	8512	0451	0204	4869	9315	3618	2294	5770
17	5241	7681	8636	4583	6453	0333	9236	5420	3231	2728	6992	0372	0505
18	8269	8466	4249	3588	6247	4845	4879	4961	4256	3328	1671	0283	5046
19	3664	0742	9329	8322	7469	8016	9345	2312	9216	9720	1558	1858	8423
20	4292	9025	8674	7876	9461	7797	4774	8519	0434	2947	4297	6655	7906
21	2482	9716	3579	9833	8203	3095	3201	6348	8936	3154	6891	0130	6843
22	6586	8113	9021	7435	9787	9556	7908	5485	7687	8669	4090	5778	1883
23	6963	9222	4465	8470	9249	1335	8562	1369	6375	6279	6268	2142	8492
24	8831	5710	6027	3773	1724	8168	6115	6682	5679	9614	0500	4508	2350
25	0960	6811	2422	3764	1629	0315	7142	9111	9381	3345	3474	3117	9086
26	3361	6816	1696	0442	0870	0703	6406	8670	6962	9612	5287	3809	9763
27	3882	9474	0174	2594	2806	0988	4621	7635	1009	9256	0131	5310	1051
28	6231	1917	1221	4241	5339	5342	2995	6003	9101	7947	1182	1139	8770
29	6614	0786	5078	0126	3759	3186	8250	3197	3603	7707	5227	4840	4092
30	6289	7457	8601	0031	2680	5920	1620	9055	8155	4750	0504	4463	6739
31	5507	9718	1984	6173	8422	0164	8501	7153	5260	3830	3447	0353	2479
32	4835	2317	4042	7768	5107	3149	7006	3980	1278	4596	6445	0784	3671
33	0807	9132	6407	1073	0861	3777	1538	0105	4441	9735	8257	7178	8905
34	2309	5153	4227	7466	6825	0689	5636	0260	4063	2244	1778	3591	6612
35	4259	4855	8242	3454	9708	7266	7494	2224	1006	8405	5852	6547	9496
36	0607	9335	0092	8420	7661	8085	4629	0608	7673	7466	1882	2875	5771
37	1236	9420	0565	1898	1083	4661	1699	1771	3463	8665	6276	4802	4898
38	4686	4409	2768	9266	9025	4690	3315	8595	6191	5287	9560	7669	5281
39	1294	9850	0535	1035	0767	9289	6569	9583	2951	7988	1181	6189	4542
40	0836	1955	3279	5438	4272	0599	4915	4794	4855	5009	0521	3291	4398
41	9744	5828	3707	9487	0271	7904	7835	5024	9735	7955	5618	6819	1242
42	3984	5109	3366	7510	0149	5370	3451	1229	1670	5582	7083	1840	7251
43	7060	1984	7477	4405	7483	5836	7099	0130	1538	5004	7973	2835	2247
45	4743	3788	8981	5815	2365	2929	1981	5400	5350	2526	4784	4528	6626
46	7317	2568	1064	4508	3706	1326	8652	1333	4440	0830	8075	6075	3667
47	5903	8293	6545	2943	9499	3422	0025	4597	9085	6940	4018	6498	2493
48	1414	7563	6736	5620	6266	3799	9797	0115	2020	0272	1554	3735	9082
49	7785	0096	5073	7156	9361	4180	8411	4435	0629	4566	1352	4577	1601
50	5022	4925	4048	5869	0389	9863	4815	2354	0042	0210	2403	2494	8104
51	2749	5504	9163	6706	7822	4106	5400	1970	4276	9047	6203	5188	2107
52	5528	3350	7255	4178	9259	1314	7476	1085	5727	9875	9850	7163	2375

4. المصطلحات الإحصائية

Acceptance sampling	معاينة القبول
Adjustment of data	تعديل البيانات
Analysis of data	تحليل البيانات
Area sample	عينة مساحية
Arithmetic mean	وسط حسابي
Auto correlated	ذاتية الارتباط
Average	متوسط
Bias	تحيز
Binomial distribution	توزيع ثنائي (ذي الحدين)
Boundaries of strata	حدود الطبقات
Census	المسح الشامل
Call□backs	إعادة الزيارة
Central limit theorem	نظرية النهاية المركزية
Cluster sampling	معاينة عنقودية
Coefficient of variation	معامل التغير
Combined ratio estimate	تقدير النسبة المركب
Conditional distribution	توزيع شرطي
Conditional expectation	توقع شرطي
Confidence interval	فترة ثقة
Confidence limits	حدا الثقة
Correction for continuity	تصحيح من أجل الاستمرار

Correlation	ارتباط
Correlation coefficient	معامل الارتباط
Correlogram	مصوّر ارتباط
Cost function	دالة التكاليف
Covariance	تغاير
Data	بيانات
Degrees of freedom	درجات الحرية
Deviation	انحراف
Difference estimation	التقدير بالفرق
Double sampling	المعاينة المزدوجة
Element	عنصر
Error	خطأ
Errors of measurement	أخطاء القياس
Estimate	تقدير
Estimator	مقدّر
Expected value	القيمة المتوقعة
Eye estimate	تقدير بالعين المجردة
Finite population correction (fpc)	تصحيح المجتمع المحدود
Frame	اطار
Frequency	تكرار
Frequency distribution	التوزيع التكراري
Frequency function	دالة التكرار
Function	دالة

Frequency histogram	مدرج تكراري
Frequency function	دالة التكرار
Geometric mean	الوسط الهندسي
Graphic presentation	تمثيل بياني
Histogram	مدرج تكراري
Hypergeometric distribution	التوزيع الهندسي الزائد
Intracluster correlation	ارتباط ضمن العنقود
Independent variable	متغير مستقل
Interpretation of data	تفسير البيانات
Interval	فترة
Interval estimation	التقدير بفترة
Inver sample	عينة معكوسة
Item	مفردة
Jackknife method	طريقة السكين الحادة
Level of significance	مستوى الدالة
Linear correlation	الارتباط الخطي
Linear regression	انحدار خطي
Linear regression estimator	مقدر انحدار خطي
mail survey	مسح بالبريد
Mean	وسط أو متوسط
Median	وسيط
New work sampling	معاينة الشبكية
Minimize	تصغير

Mode	منوال
Multistage sampling	معاينة متعددة المراحل
Neyman allocation	محاصّة نهائية
Noncoverage	عدم تغطية
Non normality	عدم الطبيعية
Nonresponse	عدم استجابة أو إجابة
Nonresponse errors	أخطاء عدم الاستجابة أو عدم الإجابة
Nonrespondent	غير مستجيب
Normal	طبيعي
Normal curve	المنحنى الطبيعي
Normal distribution	التوزيع الطبيعي
Normal standard distribution	التوزيع الطبيعي المعياري
Observation	مشاهدة
Observed frequency	التكرار المشاهد
Optimum allocation	التوزيع الامثل (محاصّة مثلى)
Ordered population	مجتمع مرتب
Ordered sample	عينة مرتب
Over estimate	تقدير بالزيادة
Parameter	معلمة
Percentage	نسبة مئوية
Periodic population	مجتمع دوري
Pilot survey	مسح أولي (اختبار الموقع)

Point estimation	تقدير نقطي
Poisson distribution	توزيع بواسون
Pooled variance	التباين المجمع
Population	مجتمع
Poststratification	تقسيم بعدْي إلى طبقات
Precision	دقة أو إحكام
Prediction	تنبؤ
Prediction equation	عرض البيانات
Presentation of data	بيانات أولية
Primary data	بيانات أولية
Primary sample	عينة أولية
Primary source	مصدر أولي
Primary unit	وحدة أولية
Probability	احتمال
Probability proportion to size	الاحتمال المتناسب مع الحجم
Proportion	النسبة
Proportional allocation	التوزيع المتناسب (محاصة تناسبية)
Qualitative characteristics	خواص نوعية
Questionnaire	استبانة
Quota sampling	المعاينة بالحصة
Random	عشوائي
Randomized response method	طريقة إجابة عشوائية

Random numbers	أعداد
Random sample	معاينة عشوائية
Random variable	متغير عشوائي
Range	مدى
Rare items	مفردات نادرة
Ratio	نسبة
Ratio estimator	مقدر النسبة
Regression	انحدار
Regression estimator	التكرار النسبي
Relative frequency	مقدر الانحدار
Relative precision	دقة نسبية
Representative	ممثل
Response errors	أخطاء الاستجابة أو الإجابة
Round number	عدد مقرب
Sample	عينة
Sample survey	مسح بالعينة
Sampling	معاينة
Sampling unit	وحدة المعاينة
Sampling without replacement	معاينة دون إرجاع
Sampling with replacement	معاينة مع الإرجاع
Sensitive question	سؤال حساس
Separate ratio estimate	تقدير النسبة المنفصل
Separate regression estimate	تقدير الانحدار المنفصل

Significant	معنوي
Simple random sampling	المعاينة العشوائية البسيطة
Size	حجم
Source of error	مصدر الخطأ
Source of variation	مصدر التغير
Square grid sample	عينة شبكة مربعة
Standard deviation	الانحراف المعياري
Standard error	الخطأ معياري
Statistical experiment	تجربة إحصائية
Strata	طبقات
Stratification	تقسيم إلى طبقات
Stratified sample	عينة طبقية
Stratum	طبقة
Student t distribution	توزيع t
Subpopulation	مجتمع جزئي
Subunit	وحدة جزئية
Supplementary data	بيانات مساعدة أو ثانوية
Survey	مسح
Systematic sampling	معاينة منتظمة
Table of random numbers	جدول الأعداد العشوائية
Target population	مجتمع الهدف
Taylor Series method	طريقة سلاسل تايلور
Test	اختبار

Three stage sampling	معاينة على ثلاث مراحل
Total variation	مجموع التغيرات
Two□phase sampling	معاينة ثنائية الوجوه
Two□stage sampling	معاينة على مرحلتين
Unaligned systematic sampling	معاينة منتظمة غير مصففة
Unbiased estimate	تقدير غير متحيز
Unit	وحدة
Variable	متغير
Variance	تباين
Variance estimate	تقدير التباين
Variance function	دالة التباين
Variation	تغير
Weight	وزن